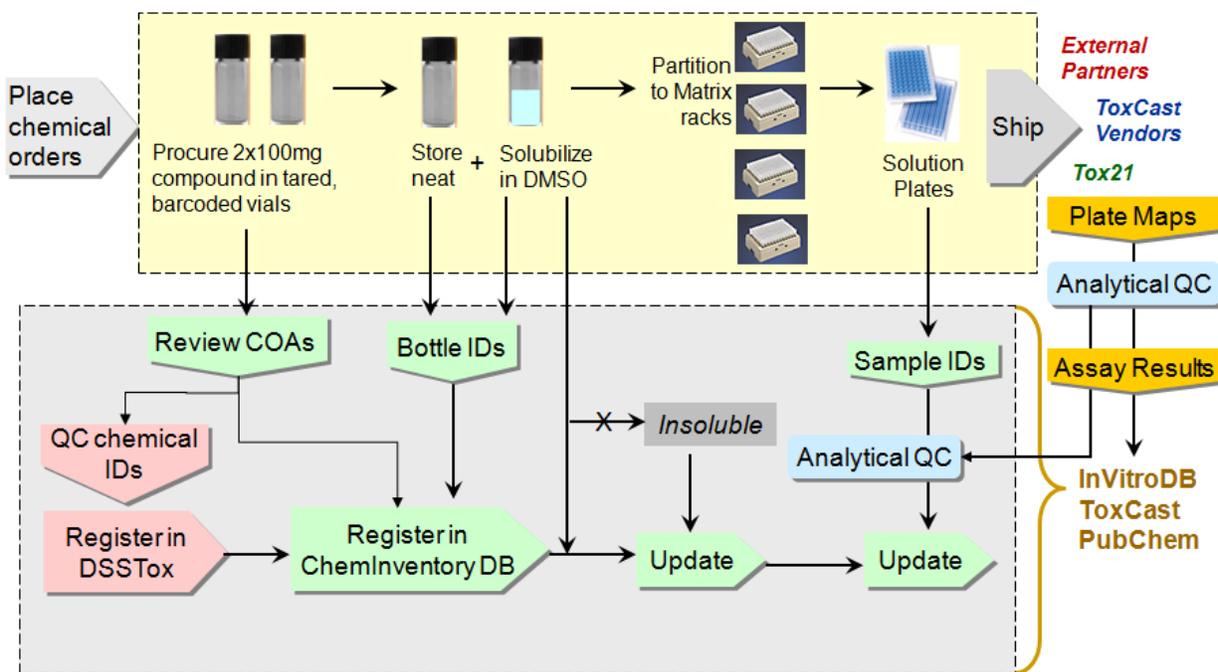


# ToxCast Chemical Inventory: Data Management & Data Quality Considerations



NOTICE: THIS DOCUMENT WAS REVIEWED BY EPA AND APPROVED FOR PUBLIC RELEASE.

12/4/2014

U.S. Environmental Protection Agency  
Office of Research & Development  
National Center for Computational Toxicology (NCCT)  
Research Triangle Park, NC 27711

## DISCLAIMER

*This document has been reviewed in accordance with U.S. Environmental Protection Agency policy. Mention of trade names or commercial products do not constitute endorsement or recommendation of use.*

## AUTHORS

Ann M. Richard, PhD ..... Principal author, ToxCast Chemical Manager & Contract Officer Representative (COR) (2007-present)

Hao Truong ..... ORISE Student Contractor to the US EPA, ChemInventory DB Developer (2012-present)

Maritja Wolf, PhD ..... Lockheed Martin Senior Scientist, Contractor to the US EPA DSSTox Project (2007-2012)

Inthirany Thillainadarajah ..... Senior Environmental Employment Program (SEEP), DSSTox Data Curator (2009-present)

## ACKNOWLEDGMENTS

The authors would like to additionally acknowledge the following persons: David Dix and Keith Houck for implementation and management of the initial ToxCast Chemical Management Contract and construction of the ToxCast Phase I\_v1 library; Robert Kavlock, Russell Thomas and Kevin Crofton for past and present leadership and support of EPA's ToxCast and Tox21 programs; Raymond Tice (NTP), Christopher Austin (NCATS), and Anton Simeonov (NCATS) for past and present leadership of their respective Tox21 programs; William Leister for heading up the Tox21 analytical QC effort at NCATS, and past and present Evotec (formerly Biofocus) Program Operations Staff - Mike Stock, Mei Steele, Kim Tran, Kim Matus, and Forum Naik.

## Table of Contents

1. BACKGROUND.....	4
1.1 ToxCast Phase I: QC lessons learned .....	6
1.2 Chemical library construction .....	7
2. CHEMICAL QC .....	13
2.1 Chemical procurements .....	14
2.2 Chemical sample management .....	16
2.2.1 Solubilizations .....	16
2.2.2 Platings.....	17
2.2.3 Shipments.....	18
2.3 Chemical information QC .....	19
2.3.1 COA Chemical Validation .....	20
2.3.2 DSSTox Chemical Information Review & Registration .....	21
2.4 Inventory data management .....	23
2.5 Sample QC .....	29
2.5.1 Analytical QC .....	29
2.5.2 Tracking sample problems .....	32
3. CONCLUSIONS – Chemical QC meeting practical and evolving needs .....	35
REFERENCES .....	37

## 1. BACKGROUND

EPA's ToxCast chemical inventory serves as the foundation of EPA's ToxCast and Tox21 research programs and has been used to generate high-throughput screening (HTS) and bioactivity data across many assay technologies and hundreds of individual assays [Dix et al., 2007; Knudsen et al., 2011; Kavlock et al., 2012; Sipes et al., 2013]. As a result, all aspects of chemical procurement, handling, data management, quality control, and structure annotation pertaining to this inventory have a direct and significant impact on the integrity and usefulness of the HTS and bioassay results generated.

EPA's National Center for Computational Toxicology (NCCT) administers all experimental and chemical handling aspects of EPA's ToxCast program through the use of extramural contract-mechanisms, which provide access to a broad range of commercial assay providers and technologies, as well as experienced high-throughput chemical sample management capabilities. The original 5 year ToxCast chemical contract was awarded in 2007 to Compound Focus Inc., a subsidiary of Biofocus DPI (South San Francisco, CA), which was acquired by Evotec in 2011. This ToxCast chemical management contract was re-competed and re-awarded for a 5 year term to Evotec in 2012 (EPA Contract No. EPD12034, <http://www.epa.gov/oam/ptod/activeindex.htm>). CFI, and later Evotec, additionally have served as the primary chemical manager for the National Institutes of Health's (NIH) Molecular Libraries Program (MLP) since its inception in 2005, creating, managing and supplying a very large chemical library (>300K), known as the Molecular Libraries Small Molecules Repository (MLSMR), to ten high-throughput screening (HTS) centers nationwide (<http://mli.nih.gov/mli/secondary-menu/mlscn/ml-small-molecule-repository/>). The NIH Chemical Genomics Center (NCGC), now a part of the National Center for Advancing Translational Sciences (NCATS), serves as both the intramural center to the MLP, as well as the main Tox21 testing facility for the multi-federal agency Tox21 program, of which NCGC and EPA are major partners, along with the National Institute of Environmental Health Sciences (NIEHS) National Toxicology Program (NTP) and the U.S. Food and Drug Association (FDA) [Collins et al., 2008; Tice et al., 2013]. NCCT's reliance on the Evotec Contract for ToxCast chemical management has directly benefitted from Evotec's broad experience in servicing these other large HTS efforts, but has required additional customization of services towards EPA's more chemical-specific programmatic needs. For the purposes of this report, we will henceforth refer to EPA's ToxCast chemical contractor, Evotec, in generic terms as "the Contractor".

EPA's chemical inventory currently consists of thousands of physical samples, including more than 4700 unique compounds, stored in powder-neat and/or solution form, along with the associated details. Sample details typically include information such as: supplier, purity, physical form, date procured, concentration, and chemical identity (chemical name, CAS – Chemical Abstracts Substance Registry Number, structure, etc.). For the purposes of this report, the terms “library” and “inventory” will be used interchangeably, and “EPA's Tox21 library” will refer only to the EPA-contributed portion of the larger Tox21 compound library, the latter consisting of more than 8K unique compounds contributed from 3 Tox21 federal partners (EPA, NTP, NCGC). Furthermore, “EPA's chemical inventory” will specifically connote all physical samples acquired and managed by EPA for possible inclusion in EPA's ToxCast testing program and/or EPA's Tox21 library, from the earliest phases of these programs to the present.

The Contract Scope of Work includes procurement and registration of chemical samples from various commercial sources (as well as storage and handling of EPA-donated samples), barcoding, weighing, and freezer storage of all samples, as well as solubilizations, dilutions, platings in various formats, shipments, and analytical quality control (QC) of solution plates, when requested. All physical samples (bottles, vials) are barcoded and an on-line electronic inventory is available to EPA that provides up-to-date tracking information on all past and present bottles/vials, including sample form (solid/liquid neat or solution), unique barcodes, supplier, lot, along with Contractor and supplier compound identifiers, date record added, quantity available, concentration of solutions, etc.

Particulars of the chemical selection criteria (e.g., function, use, regulatory interest, etc.) and enumeration of individual chemicals contained within the various inventories that comprise the different phases of testing (e.g., ToxCast Phase I, Phase II, EPA's Tox21 library, etc.) will be discussed here only in general terms and to the extent that these considerations have impacted the overall process of chemical management and review. A survey of the chemical landscape comprising ToxCast and Tox21, including details of the chemical diversity, coverage and composition of this landscape, and the factors that contributed to its construction will be published elsewhere. Finally, analytical chemistry analysis conducted over the course of the ToxCast project, primarily through partnership with the Tox21 project, will be presented in general terms as it pertains to overall library management, whereas further details and specific analytical QC results at the chemical sample (or solution) level will be published elsewhere.

The approach to chemical library construction, management and QC has been circumscribed by practical considerations (primarily cost, compound availability, and time constraints) and has been informed by and has evolved to meet the changing demands of subsequent phases of the ToxCast testing program. Given this intimate connection of the chemical management and QC process to the evolution of the ToxCast program, we provide additional historical background and context below.

### 1.1 ToxCast Phase I: QC lessons learned

There have been three major "phases" of chemical testing completed in the ToxCast program, to date. The earliest pilot phase, denoted ToxCast Phase I\_v1, tested an initial set of 309 unique chemicals (later determined to be 310), the majority pesticides. A small number of the pesticide samples were provided from the EPA Pesticidal Repository and shipped to the Contractor for subsequent processing, with the remainder of the library procured from commercial sources by the Contractor. Chemical identification relied on supplier-provided CAS and chemical names, with chemical identity and supplier-reported purity generally exceeding 98%.

Subsequently, a more thorough review of the ToxCast Phase I\_v1 chemical inventory was undertaken as part of the EPA DSSTox (Distributed Structure-Searchable Toxicity Data Network) chemical structure registration process. The EPA DSSTox project [Richard et al., 2004] has as its main goal to provide standardized annotation of accurate and consistent chemical names, CAS and chemical structures to high-interest chemical lists in environmental toxicology for use in structure-based modeling (<http://www.epa.gov/ncct/dsstox/>). This initial DSSTox review was conducted using the primary documentation provided from the EPA Pesticide Repository, and Certificates of Analyses (COAs) from commercial suppliers. *[More details on the DSSTox review and registration process applied to ToxCast are provided in Section 2.3.2 of this report.]* During the course of that review, errors in chemical names and CAS assignments were corrected, including in one case modifying a chemical name and CAS to reflect more specific stereochemistry for one of a pair of separately sourced "duplicates, yielding one fewer set of duplicates and 310 total unique chemicals (not 309) in the original set of 320 samples.

At the conclusion of ToxCast Phase I\_v1 testing, EPA employed ToxCast Chemical Contract services to perform analytical QC using high-throughput liquid chromatography mass spectroscopy (LC-MS) on a set of original Phase I\_v1 chemicals that had been solubilized at 20mM DMSO concentrations, plated and

stored under inert conditions at -20°C. These results highlighted problems and practical constraints that would inform future library and QC considerations, including:

- approximately 15% of the Phase I\_v1 library required follow-up testing with gas chromatography mass spectroscopy (GC-MS),
- a subset of approximately 25 chemicals underwent follow-up LC-MS stability testing on neat samples to confirm and account for low purity assessments,
- one of the 5 sets of separately sourced “duplicates” gave disparate analytical results (one failed identity), and
- standard LC-MS and GC-MS analytical methods were deemed inappropriate and results inconclusive for approximately 9% of the library due to low molecular weight (MW) or composition (metal-containing, mixtures).

Entering into Phase II of the ToxCast program, which would significantly expand the chemical library in both size and diversity, and also include a reprocedured version of the Phase I library (denoted Phase I\_v2), the following steps were taken:

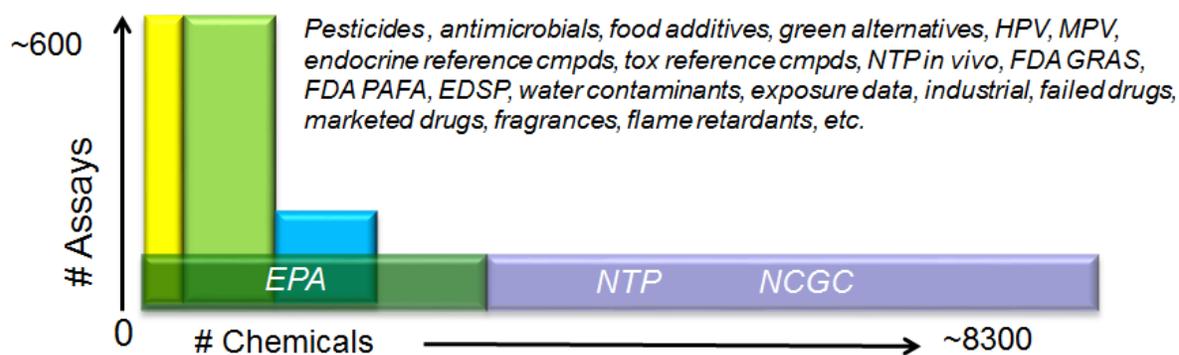
1. Removal of a class of Phase I\_v1 compounds (14 sulfurons), determined by analytical QC to undergo acid hydrolysis and significant decomposition over time in DMSO, from the reprocedured Phase I\_v2 library;
2. Review of supplier-provided information on chemical samples, in particular using information extracted from supporting documentation (e.g., COA), whenever possible, to aid in accurate chemical identification;
3. Use of a small number of preferred chemical suppliers wherever possible;
4. Procurement of 200mg samples (100mg to be solubilized to 20mM, the remaining 100mg to be stored neat for future use) to create a cost-time-efficient workflow and to minimize supplier/lot/batch variability across a multi-year testing phase;
5. Analytical QC plans for future phases of testing.

## 1.2 Chemical library construction

There are two major chemical libraries considered in this section – EPA's ToxCast library and the EPA contribution to the full Tox21 library (approximately a third of the total Tox21 library) – with the two libraries developed in tandem, and the ToxCast library fully contained within the EPA Tox21 library up to

the present. The ToxCast chemical library at the conclusion of Phase II testing consisted of a total of 1860 unique chemical substances (unique generic chemicals, with the DSSTox definition of “generic” approximately equating to the CAS level). By design, this library included the full ToxCast Phase I (v2, 293 unique compounds) and Phase II (767 unique compounds) inventories, as well as an additional set of 800 E1K compounds selected to serve long-term research objectives of the Endocrine Disruption Screening Program (EDSP21) [EPA EDSP21 Workplan, 2011]. ToxCast Phase I & II chemicals (1060 total unique) were analyzed in the full suite of ToxCast Phase II assays, whereas the additional set of 800 unique E1K compounds were run in a selected subset of these assays (approx 60) that probed various measures of endocrine activity; hence, all 1860 ToxCast compounds generated results for this subset of 60 endocrine-related assays. Additionally, all 1860 unique compounds in the ToxCast inventory were included in the initial EPA Tox21 chemical inventory (3726 compounds). [Figure 1](#) below conveys the relationships and overlaps of these various inventories thru the recent ToxCast Phase II data release.

Inventory	Chemicals	Assays	Endpoints	Completion	Available
ToxCast Phase I	 293	~600	~700	2011	Now
ToxCast Phase II	 767	~600	~700	03/2013	Now
ToxCast E1K	 800	~50	~120	03/2013	Now
Tox21-EPA	 3726	>80	>150	Ongoing	Ongoing
Tox21-Total	 ~8300	>80	>150	Ongoing	Ongoing



**Figure 1.** Listing of the Chemical x Assay dimensions and time-line for completion of the various phases of EPA’s ToxCast and Tox21 testing programs, along with an approximate indication of the overlapping coverage of the respective chemical and assay sets.

When combined with the NTP and NCGC Tox21 partner chemical inventories, the Tox21 Phase 1 inventory totaled 8307 unique substances (recently expanded to 8599 compounds with addition of a new NCGC drug plate), spanning considerable chemical structural and functional diversity. The full Tox21 inventory is undergoing HTS screening at the NCGC Tox21 robotics facility in Bethesda, MD, with a large subset of Tox21 assays focused on endocrine related endpoints and outcomes, and chemical-assay data being publicly released through PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). In addition, analytical QC for the entire Tox21 inventory, which includes ToxCast Phases I, II, and E1K, is being run in conjunction with Tox21 testing (see [Section 2.5.1](#)).

Each of the ToxCast and Tox21 testing libraries include plated replicate samples ([Table 1](#)). In this context, plate replicates are defined as solution replicates, i.e. drawn from the same stock solution created from a single supplier/lot/batch sample.

**Table 1.** Chemical library plate replicate set details and counts

Inventory		# unique cmpds in replicate set	Details	Copies per cmpd (minimum)
ToxCast	Phase I_v1	8	4 duplicates (separately sourced) / 3 triplicates	2 / 3
ToxCast	Phase I_v2	9	triplicates	3
ToxCast	Phase II	9	triplicates (minimum)	3
ToxCast	E1K	60	triplicates	3
Tox21	Phase I	88	duplicate set on each of 12 x1536 well plates, run in triplicate per assay	72

ToxCast Phase I\_v2 included a set of 9 compounds in triplicate, selected on the basis of their rich activity profiles in Phase I\_v1. Phase II incorporated the same set of 9 Phase I compounds in triplicate (minimum) for each ToxCast assay. [Note that the 9 replicates are overlapping chemicals in the Phase I and II inventories labeled within the DSSTox TOXCST structure files, available at: [http://www.epa.gov/ncct/dsstox/sdf\\_toxcst.html](http://www.epa.gov/ncct/dsstox/sdf_toxcst.html)]. The E1K compound library included a set of 60 reference compounds, selected by the EPA E1K workgroup based on known endocrine activity, plated in triplicate. Finally, Tox21 included a set of 88 compounds (containing 7 of the 9 Phase I\_v2 triplicate set)

randomly plated in duplicate on each 1536 well plate across the entire Tox21 Phase 1 testing library (12 plates total). This set of 88 unique compounds was derived from the same stock solution (and, therefore, assigned the same Tox21 ID with preface “Tox21\_4...”) and supplied by EPA to NTP and NCGC Tox21 partners. Additionally, in the initial stages of testing, each Tox21 plate was run in triplicate, with shifted plate-well configurations to optimize replicate information gain. Hence, Tox21 testing includes a minimum of 24 replicates of each of the 88 compounds across the 12 initial Tox21 plates, with 3 times that number (72) evaluated in the Tox21 assays run in triplicate.

The chronology of formation of the various chemical libraries described above strongly influenced the composition of EPA’s final ToxCast and Tox21 inventories. As previously mentioned, ToxCast Phase I compounds were primarily pesticidal active ingredients selected for testing due to the availability of extensive *in vivo* data collected by EPA’s pesticide regulatory program activities and available through EPA’s ToxRefDB [Martin et al., 2008] (<http://actor.epa.gov/toxrefdb/>). In addition, a handful of compounds of high interest to EPA programs were included, such as perfluorinated compounds, Bisphenol A, and a few known pesticide metabolites [Knudsen et al., 2011].

Beyond Phase I, significant input was solicited and received from EPA scientists and regulators, other federal Agencies, non-profits, collaborators, non-governmental organizations, and other outside stakeholders in constructing a large EPA Tox21 nomination inventory, from which EPA’s Tox21 library contribution to the full Tox21 library (consisting of contributions from all 3 Tox21 federal partners) would be constructed, and the smaller set of ToxCast Phase II compounds were to be selected. In addition, with the concurrent development of EPA’s ACToR (Aggregated Computational Toxicology Resource) database (<http://actor.epa.gov/>), hundreds of EPA and non-EPA chemical inventories pertaining to commercial use, environmental occurrence and/or of regulatory or toxicological concern were captured from public sources and cross-indexed by CAS [Judson et al., 2008; 2009].

The final nomination list of nearly 19K substances (with unique CAS) were initially assigned chemical structures from public sources available to ACToR (DSSTox, PubChem, ECOTOX, etc.) and roughly filtered by calculated physical chemistry (phys-chem) properties (using EPA’s EpiSuite, <http://www.epa.gov/opptintr/exposure/pubs/episuite.htm>) to eliminate compounds predicted to be highly volatile (low MW, high vapor pressure), unlikely to transmit through cell membranes (high log octanol/water partition coefficient - logP), or less suitable for testing and modeling (mixtures, inorganics, reactives, etc.). The resulting approximately 9K chemical names and CAS were submitted to EPA’s ToxCast Chemical Contractor for procurement, subject to the contract specifications pertaining to

cost, availability, hazard, etc. Of the final set of more than 4400 procured chemicals, approximately 8% were determined by visual inspection to be insoluble in DMSO at 20mM, or deemed otherwise unsuitable for inclusion in the final Tox21 EPA plating inventory (e.g., volatile, highly reactive).

In addition to procured chemicals, EPA entered into agreements with several outside partners to directly supply physical samples for inclusion in ToxCast Phase II (and, by association, EPA's Tox21 library).

Outside partners donating chemicals included 6 major pharmaceutical companies who donated "failed drugs" (135 total), chemical manufacturers (green plasticizer alternatives and specialty chemicals), and FDA's National Center for Toxicological Research (NCTR) laboratory (drugs with known liver toxicity).

Further ToxCast Phase II compounds were selected from the remainder of the larger EPA Tox21 inventory on the basis of their importance to EPA program office objectives, availability of *in vivo* toxicity data (EPA pesticidal programs, NTP bioassay data, FDA food additives), and known bioactivity or target interactions (drugs, reference compounds).

Prior to finalizing EPA's Tox21 inventory, another 800 compounds, including 60 reference compounds with known endocrine activity (most of which were separately procured for this purpose), were selected for inclusion in E1K testing. The majority of these compounds were already part of the EPA Tox21 inventory. Hence, the final EPA Tox21 Phase 1 library of 3726 unique substances fully incorporated the ToxCast Phase I\_v2, Phase II and E1K compound libraries, containing a total of 1860 unique compounds. The overall process of construction of this library is illustrated in [Figure 2](#) below.

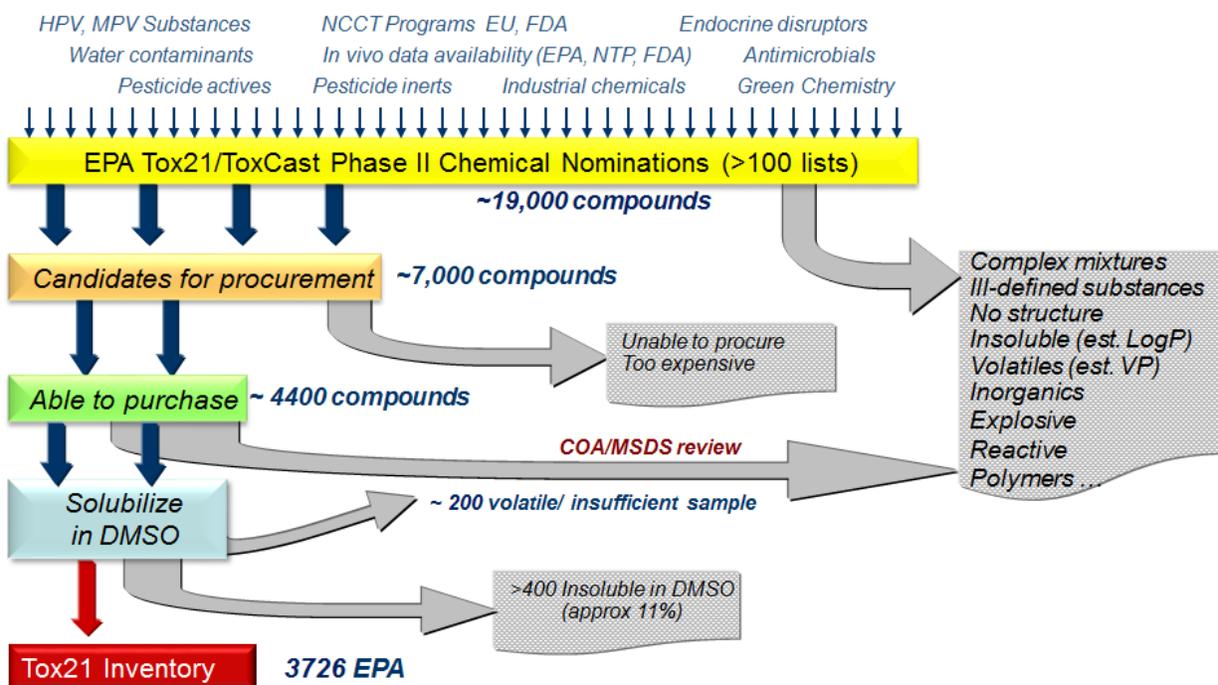


Figure 2. Schematic illustrating the main steps in the construction of the EPA ToxCast Phase II and Tox21 chemical libraries, starting with a nomination list of approximately 19K compounds, and ending with approximately 3700 procured EPA Phase I Tox21 compounds.

In the latest round of ToxCast Phase III testing, begun in 2014, the ToxCast chemical library has been expanded to incorporate additional chemicals already contained in EPA’s Tox21 Phase 1 inventory (to undergo broader ToxCast testing), as well as approximately 500 newly procured chemicals considered as part of the “EDSP universe”, and subject to many of the same filters as summarized in Figure 2. The 500 newly procured EPA chemicals, as well as hundreds of additional chemicals added by NCGC, have also been moved into what is being termed Tox21 Phase 2 testing, with the full Tox21 Phase 2 inventory, soon to be published, now exceeding 9000 unique compounds. We will make no further mention of library expansion beyond ToxCast Phase II, since the current objective is to provide details of chemical library management up to the current ToxCast Phase II data release (see <http://epa.gov/ncct/toxcast/data.html>).

## 2. CHEMICAL QC

The primary chemical QC objectives of EPA's ToxCast/Tox21 programs are twofold:

- ✓ To establish procedures to ensure the integrity and accurate tracking of chemical samples during handling, storage, solubilization, plating and transport; and
- ✓ To establish, within the practical constraints of a high-throughput testing program, the identity (CAS, name, structure), concentration, purity, and stability of the chemical samples undergoing testing.

Figure 3 provides a schematic of main elements of the chemical QC workflow, from chemical procurement through to DSSTox registration and association of chemical sample information with ToxCast/Tox21 assay results.

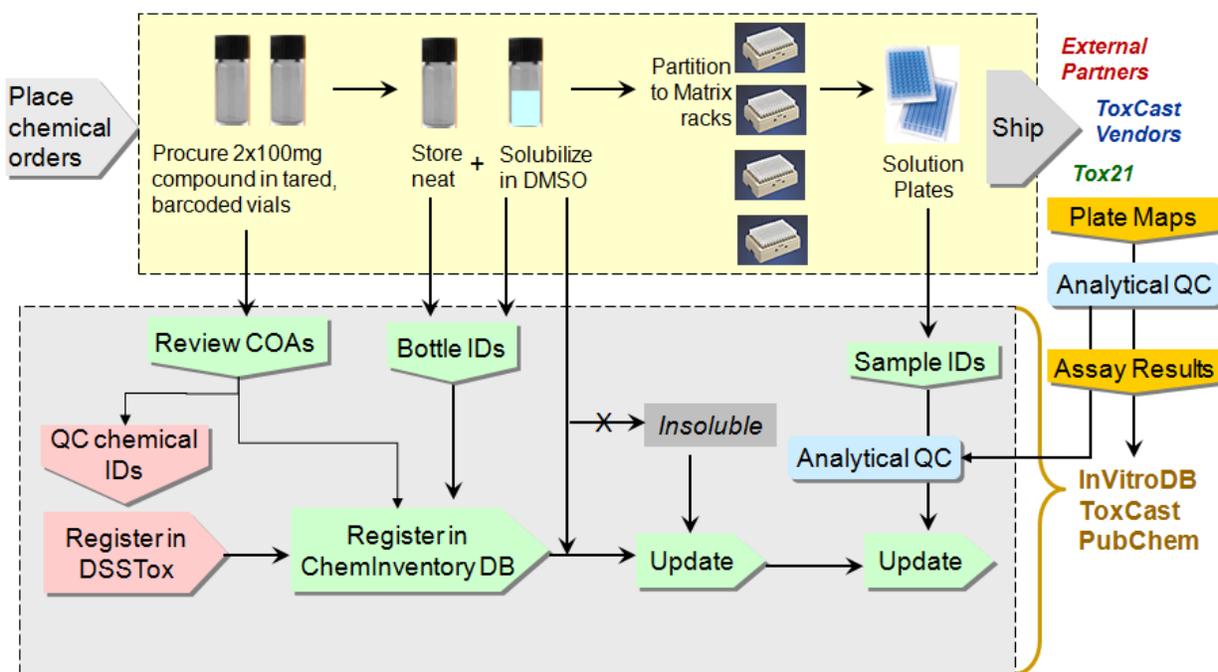


Figure 3. Schematic of the process of EPA chemical procurement through DSSTox registration, and association with ToxCast/Tox21 assay results within EPA's InVitroDB and ToxCast databases.

The EPA Inventory Report, created from the Contractor's ComIT (Compound Inventory Tracking) database and available to EPA for download as a flat MS Excel file, provides an up-to-date record of the

status of approximately 15K past and current chemical bottles or vials (neat powder or solutions). EPA's Chemical Inventory database (ChemInventory DB) was built within EPA to provide full access to a broader range of data associated with the Contractor-managed EPA inventory. ChemInventory DB is regularly updated with, and incorporates the Contractor's EPA Inventory Report content in its entirety. In addition, ChemInventory DB includes additional historical, observational, sample, and QC'd chemical structure annotations deemed necessary for support of EPA's ToxCast/Tox21 research programs. These databases provide the main information tracking resources supporting EPA's chemical inventory management.

The remainder of this section will detail current chemical QC practice within EPA's ToxCast/Tox21 programs organized into the following five subsections:

**2.1 Chemical procurements** – chemical orders, procurements and ComIT chemical registration

**2.2 Chemical sample management** – sample handling, storage, solubilizations, platings, shipments

**2.3 Chemical information QC** – chemical sample information review and DSSTox registration.

**2.4 Inventory data management** – data information management and sample tracking within ChemInventory DB

**2.5 Sample QC** – analytical chemistry QC analysis, tracking observed problems with physical samples, etc.

## 2.1 Chemical procurements

EPA's placement of a chemical procurement order involves issuance of a Task Order and providing the Contractor with a list of generic CAS and chemical names for procurement; SMILES may also be provided and are used by the Contractor for structure searching of commercial sources through larger aggregated services, such as eMolecules (<https://www.emolecules.com/>) or ChemNavigator (<http://www.chemnavigator.com/>). After supplier-quotes are received, these are reviewed by the EPA Contract Officer Representative (COR) for adherence to pricing and quantity requirements, and the final orders, usually requiring multiple suppliers, are placed. The process of chemical procurement is by far the most time-consuming step in the ToxCast chemical management workflow, with the time between placements of orders, receipt and processing of quotes, and receipt of samples usually spanning weeks to months for delivery (with delays based on multiple shipments from various suppliers, backorders, or

advertised chemicals being unavailable once the order is processed). In addition, the chemicals ultimately received and the information provided by suppliers with shipments often is incomplete (e.g., missing CAS, names, or inaccurate associations of CAS-name-structures) and/or does not match the CAS and name listing provided with the original order. In these cases, further chemical information review is conducted to reconcile original orders with received chemicals and accurately annotate newly procured substances. We refer to this information review step, henceforth, as “chemical data validation”.

The standard process of chemical procurement, which applied to the majority of samples incorporated into EPA’s ToxCast and Tox21 inventories, involves the purchase of 2 identical bottles of 100mg neat (dry) or 10 ml pure liquid, analytical grade samples (> 98% purity). A company may be designated a “preferred supplier” based on historical reliability in providing samples and requested documentation, adhering to promised shipment schedules, providing a sufficiently large catalog of chemicals, and willingness to repackage samples in desired quantities. The majority of samples were shipped directly to the Contractor in pre-tared, barcoded vials provided by the Contractor for this purpose. The decision to order by standardized weights (100mg) rather than molar quantities (MW-dependent) was to streamline and optimize the efficiency of the procurement process, whereby one of the two 100mg bottles is designated for solubilization (without need for a more costly and time-consuming manual weighing step) and the other stored in neat-powder form. The 100mg quantity typically yields 10-20 ml of 20mM DMSO solution for a compound in the intermediate 250-500 MW range [*Note that the total amount of 20mM solution consumed through the entire ToxCast Phase I\_v1 testing per compound was on the order of 8 ml*]. Upon receipt of procured chemicals by the Contractor, bottles are scanned, weighed, registered into the Contractor ComIT database, and either immediately solubilized, or stored in a -20°C freezer under inert conditions until a solubilization order is placed

EPA requires, wherever possible, a COA and MSDS (Material Safety Data Sheet) from the supplier, which the Contractor provides to EPA in pdf electronic form. Samples are accepted by EPA without documentation when a supplier cannot be located otherwise, or when the supplier does not provide the requested documentation after repeated requests (COAs are missing for approx. 15% of unique lot-batch chemical samples). The receipt of such documentation often lags the receipt of the chemical samples by weeks to months, causing further delays in processing. The supplier is requested to also provide a chemical structure file and MW for defined pure compounds (SMILES or SDF file). The Contractor's ComIT database records the Supplier, Catalog number, Lot number, the Contractor shipment sample code, SMILES, molecular weight (usually derived from the structure file), physical form

of the chemical received (solid or liquid), quantity (ul or mg), and the date the sample was registered in ComIT.

In cases where a chemical name, CAS or structure are not provided by the supplier, the Contractor attempts to fill in this information through reference to the original orders or supplier website catalogs, or through internal database matches to the compound structures provided. Most common deviations from this standard procurement and registration process for a subset of the chemicals (less than 10% of Phase II and Tox21 chemicals, whereas a greater percentage of reprocured ToxCast Phase I and Phase III chemicals) involves use of smaller, specialty suppliers and procurement of larger or smaller quantities for hard-to-locate chemicals, with samples shipped in prepackaged, supplier-provided containers, requiring barcoding, splitting, weighing and transfer to the Contractor barcoded vials prior to solubilization. At various stages during the procurement process, or at the conclusion of a procurement order, new bottles are received, barcoded and registered in the ComIT system, with the information available on demand to EPA through the on-line EPA Inventory Report.

## 2.2 Chemical sample management

The EPA Inventory Report provided by the Contractor's ComIT tracking system provides a window into the physical nature of EPA's chemical library and the various processes necessary for creating and maintaining that library. These include sample handling, storage, weighing, solubilizations, platings, shipments. Standard Operating Procedures (SOPs) and industry guidelines are adhered to in the general handling, weighing and storage of chemical samples associated with EPA's chemical inventory. Neat samples are weighed under a hood, and all DMSO solubilizations, dilutions, solution transfers, and platings are carried out under inert atmosphere (Nitrogen) conditions to minimize moisture uptake. All samples are subsequently stored sealed in barcoded supplier containers, or pre-tared vials (after weighing and transfer from supplier containers) in a -20°C walk-in freezer under inert conditions.

**2.2.1 Solubilizations:** When samples are procured in standard 2 x 100mg tared vial format, the standard process of solubilizations involved fully solubilizing one of the pair of 100mg neat sample bottles in DMSO to create a stock solution at the desired target concentration. This process, applied to the major portion of the EPA inventory, avoided the need for a separate manual weighing step for processing thousands of samples, a step which is costly and labor intensive, and likely more error-prone. In some cases, for particularly potent or high priority chemicals, the target concentration was allowed to be as low as 5mM. For most chemicals tested through ToxCast Phase II, however, chemicals not soluble at

20mM in DMSO were labeled as “Insoluble” and excluded from testing. Approximately 8-10% of the newly procured chemicals comprising the Tox21 library (Phase II and beyond) were deemed insoluble at 20mM in DMSO.

Since 2013, there has been increased demand for broader chemical coverage (i.e., including previously insoluble chemicals), as well as higher concentration platings, up to a maximum target concentration of 100mM in some cases (to reduce the relative concentration of DMSO). Hence, all previously labeled “Insoluble” chemicals, as well as all new procurements, are now solubilized to a “maximum achievable concentration” (MAC) up to 100mM, or if insoluble at that top concentration, tested in 4-6 increments down to a minimum of 5mM. Approximately 20% of previously insoluble chemicals at 20mM were later judged to be soluble at lower concentrations of 10mM or 5mM in DMSO and, thus, became candidates for inclusion in Phase III testing. In addition, working stores of 20mM solution are drawn off of the higher concentration stocks and are available to fill 20mM plating orders. For large MW compounds, polymers, or ill-defined substances with unknown MW, a weight-based stock solution concentration of 10mg/ml DMSO is used as the equivalent 20mM target concentration (using an average MW of 500 g/mol, this equates to 20mM). Solubilization results are reported to EPA, with each bar-coded solution vial assigned the designation “Soluble” or “Insoluble” within ComIT, and accompanied by a solution concentration and total quantity (ul) of solution available. Samples labeled “Insoluble” are retained and stored in the freezer with the remainder of the library.

**2.2.2 Platings:** Plating orders to be shipped to ToxCast assay vendors, EPA collaborators, Tox21 partners, or MTA (Material Transfer Agreement) partners are submitted by NCCT staff to the COR with the following requested information: inventory (e.g., ToxCast Phase I\_v2) or a list of compound CAS and names, compound sample volume (typically 50-100ul), target concentration (20 or 100mM), and plate details or special requests (e.g., 96 or 384 well plates, round or V bottom, and custom compound plating configuration – default is 96 well, V bottom, with all wells filled in no particular order). The COR checks the status of available sample stocks for the requested compounds in ChemInventory DB to ensure that sufficient quantity of solution stock is available at the requested target concentration and, where possible, may attempt to match specific supplier-lot-batch details to previous test plates. If sufficient solution at the target concentration is unavailable to fill the plating order, the order may trigger additional orders for solubilizations (of available neat sample), dilutions (of 100mM solution stock to 20mM stock), or reprocurements (if sample stock is depleted or degraded).

The EPA COR supplies the Contractor with a list of solution vial barcodes, and specifies plate details, solution volume, target concentration, shipping address and contact. In most cases, a particular sample-well plating order and inclusion of randomly placed sample replicate solutions are additionally specified by the COR. All Contractor plates are barcoded and the process of plating is fully automated and accompanied by final dilutions to the requested target concentration and precise determination of plated solution concentrations and volumes. Plates are heat sealed, frozen and shipped overnight, well-packed on dry ice. Serial dilutions usually accompanying assay runs are subsequently carried out by ToxCast and Tox21 vendors and collaborators. In addition, once delivered, plates are stored, handled and disposed by each vendor and collaborator according to standard practice within their laboratory, although it is recommended that plates be stored covered and frozen at -20 degrees C when not in active use.

**2.2.3 Shipments:** At the time of shipment (usually within 1-2 days of plating), the Contractor provides the EPA COR with an electronic plate map file generated from the Contractor's ComIT database (csv format). This file lists Contractor and supplier-provided compound sample identifiers, the plate barcode(s) and well addresses of each sample, the final concentration and volume of each plated sample, the date of plate creation, and linkage of the plated solution barcode to a parent barcode if the former was generated in the course of plating and was not previously in ComIT. In addition, EPA maintains a record within ChemInventory DB of every Contractor shipment, along with the shipment number, recipient (vendor or collaborator, address, contact info), inventory (e.g., ToxCast PhI\_v2, etc.), and plate barcodes associated with the shipment (which link to plate map details within the database).

At the time of a plate shipment, the EPA COR processes the Contractor-generated plate map file, strips the file of chemical identifiers, and replaces these identifiers with an EPA\_Sample\_ID. EPA's standard protocol with all ToxCast assay contractors and collaborators is to provide blinded plate maps during the assay testing stage, with chemical identities unmasked only after the assay data has been delivered to EPA. The EPA\_Sample\_ID historically was linked to the registered Contractor solution barcode, with distinct EPA\_Sample\_IDs generated for the randomly placed sample replicates to mask their identities. A new protocol recently has been instituted in which EPA\_Sample\_IDs are auto-generated from the unique plate barcode and well address of each shipment, to mask compound identities as well as to link samples unambiguously to plated concentrations. ChemInventory DB maintains a record of every assigned EPA\_Sample\_ID provided to assay contractors and collaborators, mapped to the original solution barcode. The blinded plate map file is provided by EPA, usually within a day of the plate

shipment, to the assay vendor or collaborator. All assay results subsequently generated by assay vendors or collaborators, in turn, are reported back to EPA by EPA\_Sample\_ID, which is linked through the vendor shipment back to the original plate barcodes and concentrations, and is mapped to DSSTox chemical identifiers through ChemInventory DB tables, the latter now fully incorporated into EPA's ToxCast assay data pipeline.

### 2.3 Chemical information QC

As alluded to above, during the course of chemical information review associated with chemical procurements, a number of problems were frequently encountered. These included:

1. Incomplete supplier-provided information in the Contractor's ComIT database (e.g., missing chemical name and/or CAS from supplier, or missing structure).
2. Supplier-provided chemical identifiers (CAS, name) that were missing important substance details (stereochemistry, salt or hydrate form, etc.) or did not agree with the information extracted from the supplier-provided COA or MSDS information.
3. Conflicting CAS and chemical name information (e.g., CAS is invalid, or CAS and name are mismatched and do not correspond to the same substance).
4. Conflicting CAS-name to structure associations, with structures provided by suppliers (or EPA's Contractor) missing salt, complex or stereo information.
5. Completely wrong CAS and chemical name provided by the supplier for the procured sample (i.e., a different chemical entirely).

During the early phases of ToxCast (Phase I\_v1 and initial compilation phases of Phase II and Tox21), the frequency with which incomplete information was provided by chemical suppliers (10-20%) and errors in name-CAS-structure associations were detected (5-10%), and the occasional major errors associated with the primary chemical identifiers (less than 1%), led EPA to institute additional chemical information QC for all future chemical procurements. This review consists of two major steps: 1) "COA chemical validation" to establish chemical identity (chemical name, CAS, MW) from the supplier-provided COA; and 2) DSSTox chemical information review to ensure accurate and consistent substance (CAS, name, description) and structure annotation of the generic chemical (independent of supplier, lot, batch) as part of the DSSTox chemical registration process.

**2.3.1 COA Chemical Validation:** Once chemicals are barcoded and registered by the Contractor in ComIT, and the requested COA and MSDS documentation have been received, the latter documents are subject to review and data extraction. This process was originally performed by EPA staff through Phase II procurements, whereas these services have been substantially shifted to the Contractor since 2013. Once received by Evotec, COA and MSDS files are renamed according to a standard convention linked to the original bottle barcode (e.g., TX000145\_COA.pdf, where TX000145 is the bottle code corresponding to a physical sample registered in ComIT). Next, information is extracted from the COA/MSDS pdf files that pertains to chemical identity (name, CAS, hydration, salt form, molecular weight, density) and method of analysis (analytical method, purity determination, expiration date), along with safety cautions extracted from the MSDS. Due the varied non-standard formats and scanned quality of received COAs and MSDSs, and the corresponding difficulty of auto-text extraction or optical text recognition (OCR), this information requires oversight and some manual data entry. The Contractor currently employs a combination of text-extraction scripts followed by manual review and data entry, where necessary. Data is entered into a standardized COA table format, which is provided to EPA in electronic tabular form in association with chemical procurements, and prior to invoicing. Three sample entries from the current COA table (transposed) are provided in [Table 2](#) below.

**Table 2.** Sample entries from the Contractor-provided COA table resulting from the chemical data validation process, with the last 3 fields added by EPA following DSSTox chemical review and registration.

Bottle_Barcode	TX013709	TX0012899	TX014193
COA_FileName	TX013709_COA_MSDS.pdf	TX0012899_COA_MSDS.pdf	TX014193_COA_MSDS.pdf
Data_Extraction_Status	Success!	Success!	MSDS not available
COA_Product_No	MKBP4248V	12079	40391
COA_Lot Number		A0308579	20130220
COA_ChemicalName	1, 3-Butanediol dimethacrylate-contains 150-250 ppm MEHQ as inhibitor, 95%	Hexyl alcohol	Argatroban monohydrate
COA_CAS	1189-08-8	111-27-3	141396-28-3
COA_MolecularWeight	226.27		526.65
COA_Density	1.01		
COA_Purity_(%)	94.20%	98.8%	98.8
COA_Methods	GC	GC	HPLC
CoA Test Date	5/29/2013		3/11/2013
COA_ExpirationDate			2/1/2015
MSDS_Cautions	May be harmful if inhaled. Causes respiratory tract irritation. May be harmful if swallowed. May be harmful if absorbed through skin. Causes skin irritation. Causes eye irritation.	Flammable liquid and vapor. Harmful if swallowed. Irritating to eyes and skin	
COA_GSID_Mapping	complete	complete	complete
COA_ReviewNotes	CAS-name-GSID checked	CAS-name-GSID checked	Parent in DSSTox, added monohydrate
DSSTox_GSID	44784	21931	57888

In cases where either or both the COA and MSDS are missing, the COA chemical validation step will rely upon whatever supplier-provided information is available, or additional information on chemical identity may be located on supplier and manufacturer websites. In a small number of cases, when a chemical annotation is corrected, the MW can substantially change, and with it the reported solution concentration that was based on the original MW, triggering EPA adjustments to concentrations associated with plated chemicals.

**2.3.2 DSSTox Chemical Information Review & Registration:** Once the chemical identity of a sample has been established to the extent possible from supplied documentation, further review takes place within the EPA DSSTox project to ensure consistency and accuracy of assigned DSSTox substance (CAS,

chemical name) and structure annotations. The last 3 fields in the COA table above are added by the EPA reviewer and pertain to the DSSTox review and registration process during which a final DSSTox\_GSID (generic substance ID) is assigned. The DSSTox\_GSID links the chemical bottle (Bottle\_Barcode) and all derived samples (neat, stock solutions, daughter solutions, etc.) to the DSSTox generic chemical identifiers (CAS, name) and chemical structure within the DSSTox database (DSSTox DB).

The DSSTox project is recognized for the high level of QC review applied to the registered content, providing accurate associations and, wherever possible, unique mappings of CAS-chemical name to DSSTox\_GSID and to chemical structure. Non-unique mappings (e.g., 2 GSIDs assigned to a single structure) historically only occurred with the assignment of DSSTox\_GSID to a “representative structure”, and were accompanied by annotations clarifying the nature of these approximate substance to structure associations. In addition, salts, stoichiometric complexes (including hydrates), and stereochemistry (geometric – E/Z, and chiral – R/S) are explicitly annotated within a DSSTox chemical record and are assigned a unique DSSTox\_GSID. In the context of the ToxCast and Tox21 testing efforts, this chemical detail is captured to the extent that this information is communicated or available from a publication or chemical supplier. More details on DSSTox chemical information review procedures can be found at: <http://www.epa.gov/ncct/dsstox/ChemicalInfQAProcedures.html>. More information on DSSTox Standard Chemical Fields, applied across all DSSTox published chemical inventory files, can be found at: <http://www.epa.gov/ncct/dsstox/MoreonStandardChemFields.html>. [*Note, updated DSSTox Standard Chemical Fields within the current ToxCast chemical files associated with the Phase II data release have slightly modified and truncated from their original form.*]

The nature and frequency of CAS-name-structure errors encountered in past DSSTox curation efforts applied to published chemical lists (see <http://www.epa.gov/ncct/dsstox/DataFiles.html>) are consistent with those encountered during the course of the EPA ToxCast and Tox21 projects in processing of information provided by various chemical suppliers. These errors are reduced but not eliminated even with the additional COA chemical validation step. In particular, errors in CAS-name associations are not uncommon, with deleted and invalid CAS, as well as mis-matched CAS-name assignments encountered. In addition, chemical structures associated with CAS-name information in the public domain, and by chemical supplier websites, can be of insufficient precision or incorrectly assigned to the stated CAS-name of the procured substances. Most often, these errors are relatively “minor” and of 3 general types: salt-parent compounds not accurately distinguished (e.g., a parent structure provided for a salt

CAS-name, or vice versa), explicit complexed moieties or waters (hydrate) not accounted for either in the CAS-name or in the structure (but specified in the COA), and missing or inadequately represented stereochemistry (e.g., specified as E-form in CAS and/or name, but listed as Z form or unspecified in the chemical structure). The final EPA review of the COA table information following the COA chemical validation step, and mapping of each new bottle barcode to a registered DSSTox\_GSID, or creation of a new DSSTox\_GSID record, is required to complete a new data entry into the ChemInventory DB.

Lastly, although EPA did not procure or manage the plating of Tox21 chemical samples from the non-EPA Tox21 partners (NTP and NCGC), EPA performed a substantial amount of chemical information QC on the NCGC supplier-provided samples, and both the NCGC and NTP Tox21 chemical inventories were subject to the standard DSSTox chemical annotation review procedures and assigned to DSSTox chemical structures. Hence, all Tox21 substances are registered in DSSTox, which is the source for chemical substance and structure annotations for the entire Tox21 inventory. In addition, the mapping of unique Tox21 stock solution IDs (sample ids), used for reporting of Tox21 assay results in PubChem, to DSSTox substance identifiers (DSSTox\_GSIDs), is centrally stored and tracked within EPA's ChemInventory DB.

## 2.4 Inventory data management

As indicated earlier, chemical inventory data management currently has two major components: 1) the Contractor ComIT internal tracking database, from which an up-to-date EPA Inventory Report (Excel file) can be dynamically generated at any time by the EPA COR through a secure website; and 2) EPA's ChemInventory DB that fully incorporates the EPA Inventory Report, along with EPA-added content pertaining to sample details (including COA and MSDS extracted information on purity, analysis method, cautions, etc.), platings, and shipments, as well as a DSSTox GSID that links to QC'd chemical identifier and structure information. [Table 3](#) below lists the typical fields contained within the ComIT Excel file.

**Table 3.** List of data fields contained within the current ComIT EPA Inventory Report, along with a brief description of the field contents.

ComIT EPA Inventory Report (05/15/2014)	
Field	Description
Barcode_Parent	Parent bottle barcode when samples are received in Supplier containers
BARCODE	Primary key - unique bottle barcode ID used as EPA_Sample_ID in most cases
STATUS	Status of bottle: Available, Disposed, Shipped
COMPOUND_NAME	Supplier-provided chemical name (or, if missing, may be retrieved from supplier website or EPA order)
CAS	Supplier-provided CAS (or, if missing, may be retrieved from supplier website or EPA order)
VENDOR	Supplier
VENDOR_PART_NUMBER	Supplier part number or catalog number
QTY_AVAILABLE_MG	numeric entry only if sample is in neat or powder form (mg)
QTY_AVAILABLE_uL	numeric entry only if sample is solubilized (ul)
CONCENTRATION_mM	concentration in DMSO only if QTY_AVAILABLE_ul entry
QTY_AVAILABLE_UMOLS	convert quantity (mg or ul) to umols based on reported MW
STRUCTURE_REAL_AMW	Molecular Weight calculated from the structure
SAM	Contractor sample ID, unique to shipment/supplier/compound
CPD	Contractor compound ID, unique to assigned structure
PO_NUMBER	Contractor PO number
LOT_NUMBER	Supplier-provided sample lot (batch)
FORM	SOLID, LIQUID, SOLUTION
Date_Record_Added	Date bottle or vial BARCODE added
SOLUBILITY_DMSO	Soluble, Insoluble, "blank" (if neat)
SOLUBILITY_DETAILS	Solubility observations (cloudy, colored, etc.) – new field

A snapshot of the actual content of the Contractor-generated EPA Inventory Report, as of 5/15/2014 is as follows:

- 14945 bottle barcode entries, including all historical entries and empty containers
- 13851 bottle entries with available sample:
  - Approx. half are neat-mg, the remainder are solutions-ul

- 4560 unique names and 4676 unique CAS for available samples
- 4946 unique structures for available samples
- 1149 bottles (8% of total) missing a supplier name (865), CAS (284), or both (121)
- Available samples from 31 commercial chemical suppliers, or provided by EPA (fewer than 200 chemicals):
  - 58% of samples from a single major chemical supplier
  - 87% of samples from the top 3 chemical suppliers
  - 13% difficult-to-procure chemicals obtained from 28 smaller chemical suppliers

In addition to fully incorporating the above ComIT content, the following information is currently tracked within ChemInventory DB (as of 5/15/2014):

- > 7700 COA-MSDS (or either COA or MSDS) files and associated, extracted content in COA table
- > 40 assay vendors or collaborators – address & contact info for recipients of plate shipments
- > 20K unique EPA\_Sample\_IDs assigned, i.e., includes IDs for replicates and > 12K Tox21\_IDs
- > 150 plate shipments to date
- > 900 unique plate barcodes shipped
- > 90K plate wells filled, with each well address linked to EPA\_SAMPLE\_ID, volume (ul) and concentration (mM) information
- > 8K unique LotMatch\_IDs, constructed within ChemInventory DB to link common sets of bottles (neat and solution) with matched compound, supplier and lot-batch.
- > 5K unique DSSTox\_GSIDs assigned across the entire EPA inventory, and > 9K if the full Tox21 inventory is considered

Information pertaining to all Contractor-managed aspects of EPA's chemical inventory flows to EPA through the on-line ComIT-generated EPA Inventory Report, along with separate Contractor-generated electronic reports delivered to EPA in association with completion of procurements, solubilizations, platings, and shipments. Prior to 2012, the bulk of this information was stored in Excel files, and portions were managed within an MS ACCESS database. In early 2013, all information tables and files were incorporated into a single MySQL database, "ChemInventory DB", built within NCCT for the purpose of consolidating and automating chemical management duties. Concurrently, several EPA task orders were issued to the Contractor to expand content of the ComIT EPA Inventory Report to provide EPA with readily available information for assessing sample status (solution, neat), solubility (soluble, insoluble), and availability (quantity available – mg, ul, mmols), as well as to standardize the data

format, to the extent possible, of reports provided to EPA so as to facilitate auto-processing data entry and EPA placement of new procurement, solubilization, and plating orders. The expanded ComIT-generated EPA Inventory Report, along with ChemInventory DB, have significantly improved the efficiency, quality and integrity of EPA's chemical data management, while providing greater access to database information through automated queries (e.g., to generate unblinded plate maps) and enabling direct linkage to the ToxCast assay data processing pipeline. Figure 4 below conveys the level of detail and complexity of the MySQL data model captured within the current ChemInventory DB.

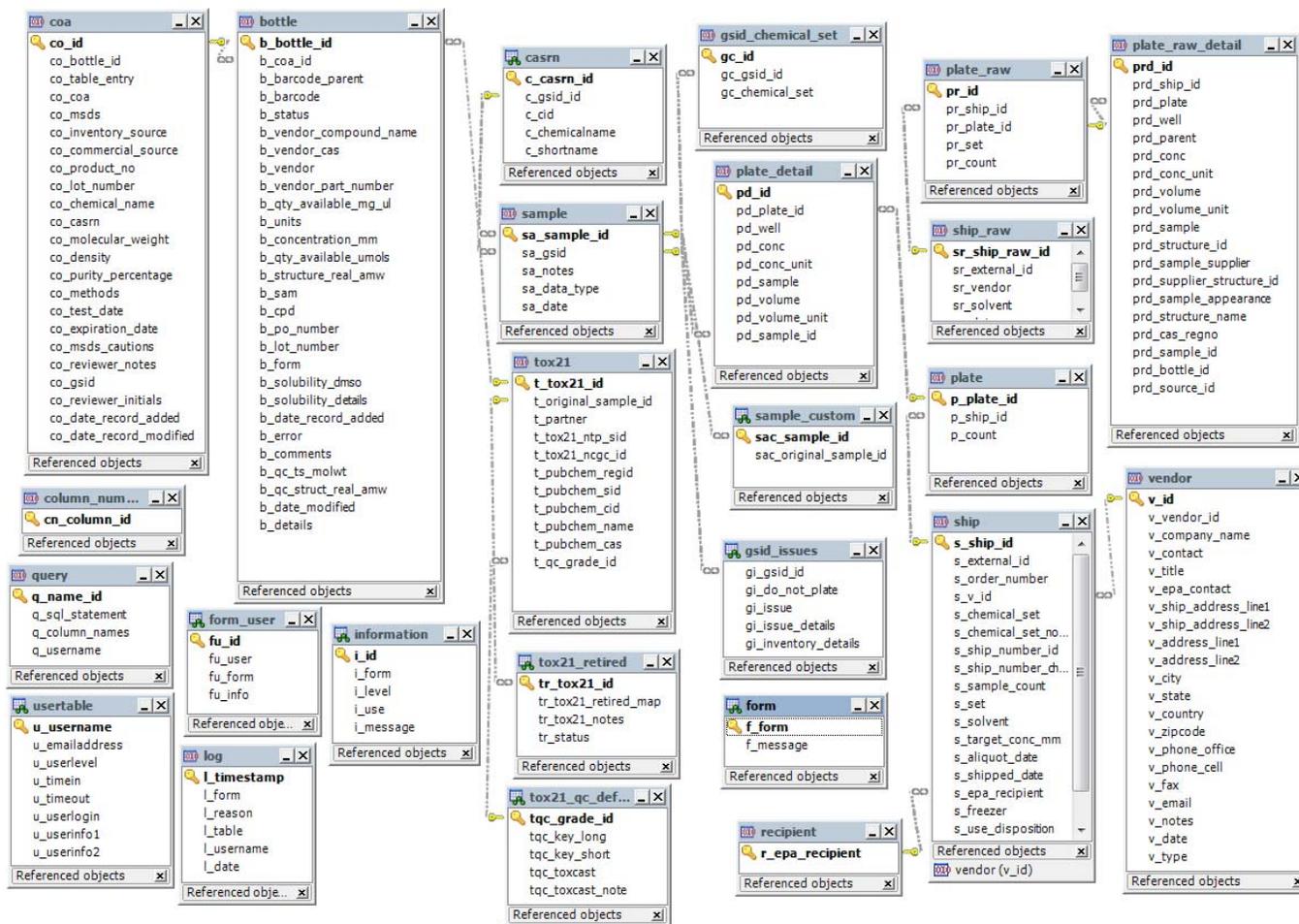


Figure 4. ChemInventory DB data model relationship schematic (as of 12/4/2014).

The DSSTox chemical review and registration described in Section 2.3.2 is separately applied to every sample in ChemInventory DB, either prior to or concurrent with placement and processing of chemical

procurement orders. The DSSTox database spans a large number of public chemical inventories outside of ToxCast and Tox21, and is a separately maintained database from ChemInventory DB. Work is currently underway within NCCT to dynamically link ChemInventory DB to the DSSTox DB through the DSSTox\_GSID to allow ChemInventory DB to access the most current DSSTox chemical information available. This relationship and the relative sizes of information components across the 2 databases are represented in Figure 5 below.

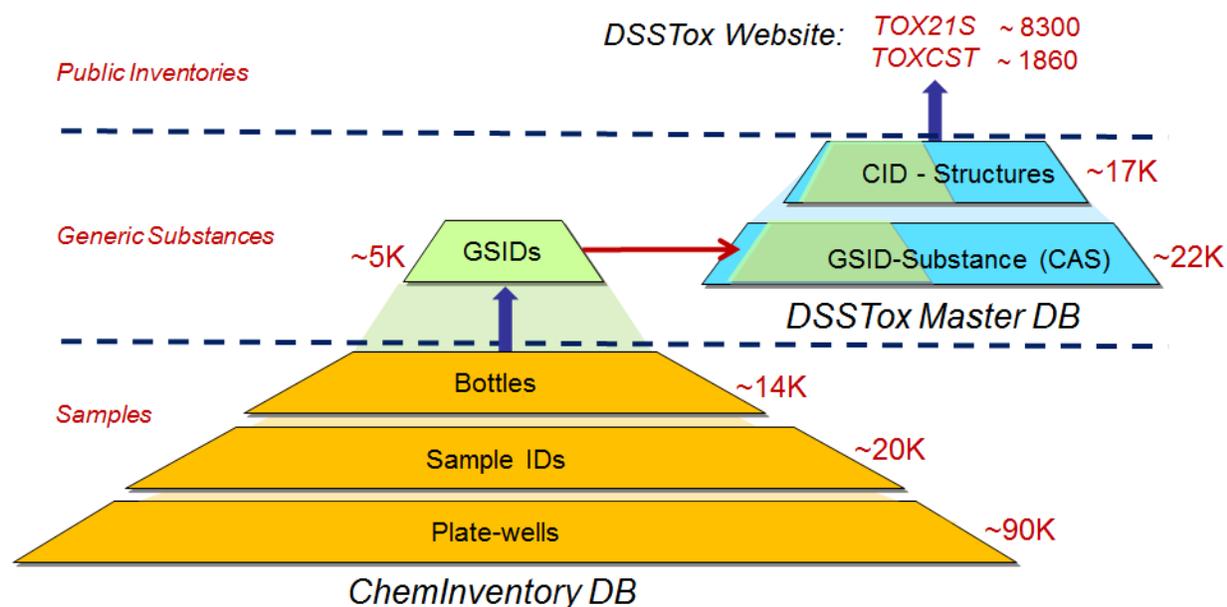


Figure 5. Schematic illustrating the relationships of components of ChemInventory DB relative to the DSSTox Master DB and public DSSTox inventories, TOXCST and TOX21S (as of 5/14/2014).

The generic chemical component of the plated ToxCast and Tox21 chemical inventories are represented as “Inventories” within the DSSTox DB, as well as published as separate DSSTox Data Files on the public DSSTox website (TOXCST and TOX21S, respectively). DSSTox Inventories contain the unique listing of DSSTox\_GSID substances, along with associated chemical and structure fields. The respective SDF Download Pages can be found at: [http://www.epa.gov/ncct/dsstox/sdf\\_toxcst.html](http://www.epa.gov/ncct/dsstox/sdf_toxcst.html) and [http://www.epa.gov/ncct/dsstox/sdf\\_tox21s.html](http://www.epa.gov/ncct/dsstox/sdf_tox21s.html).

In summary, EPA's ChemInventory DB consolidates all QC'd chemical information pertaining to EPA's ToxCast and Tox21 chemical libraries (including tracking the NTP and NCGC chemical stock solution IDs and source IDs) in association with all plated solutions submitted for testing. Assay results are linked to a shipment and plate details, including EPA Sample IDs or Tox21 solution IDs, which in turn are linked to generic chemical identifiers (through DSSTox\_GSIDs) within ChemInventory DB. File exports are provided to Tox21 partners, whereas ChemInventory DB data tables can be directly accessed within NCCT's ToxCast data pipeline to support EPA's ToxCast and Tox21 HTS programs and data analysis. The central role of ChemInventory DB to the entire process of chemical management is schematically illustrated in Figure 6 below.

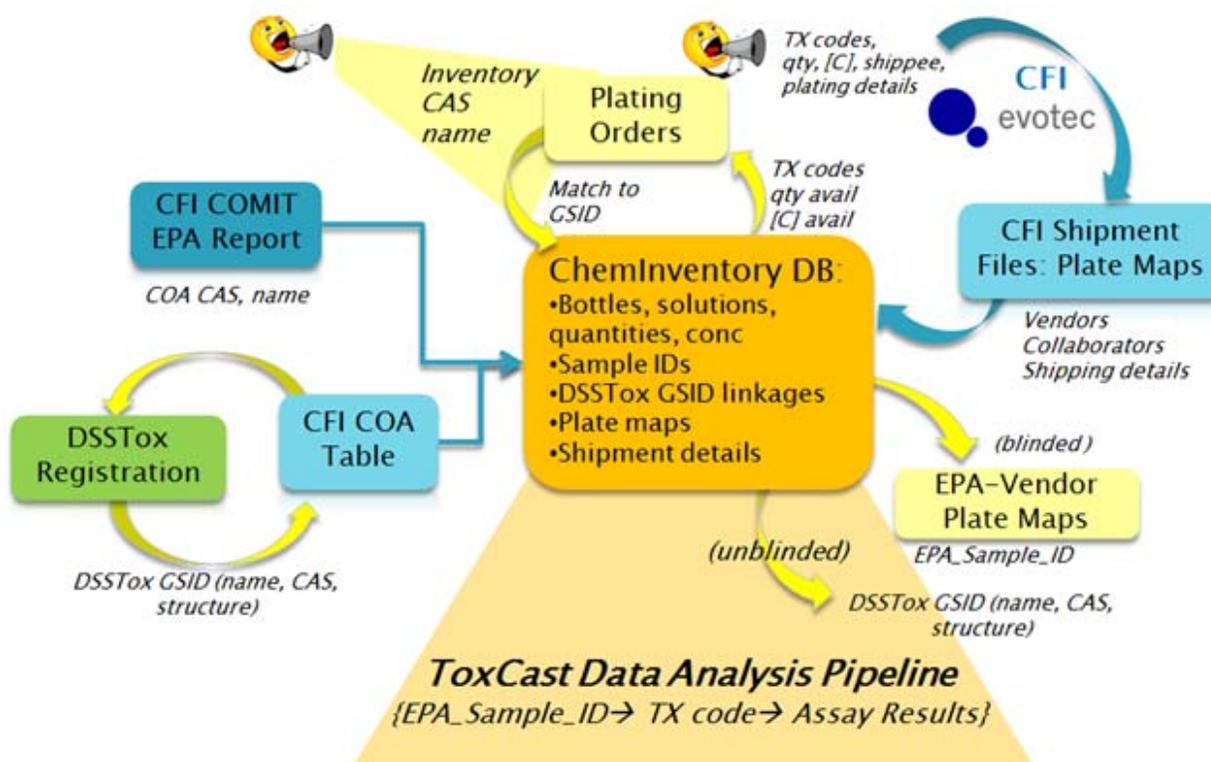


Figure 6. EPA's chemical management processes centrally linked to ChemInventory DB.

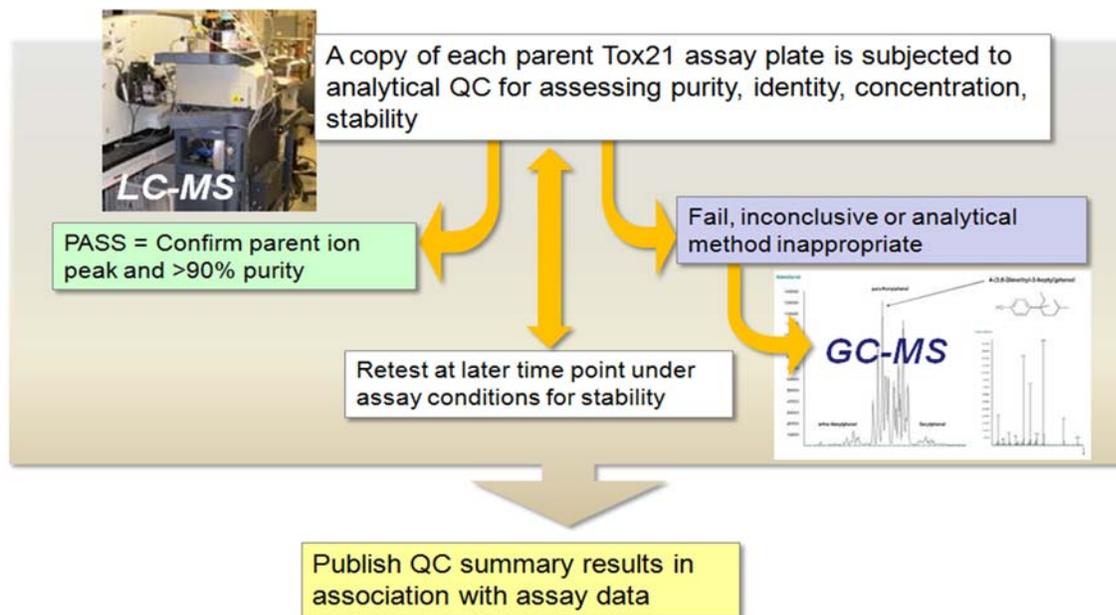
## 2.5 Sample QC

Most of the previous discussion has focused on chemical information QC pertaining to establishing the identity of a tested sample with respect to accurate associations of CAS, chemical name, and chemical structure. Significant emphasis has been placed on this type of QC within ToxCast and Tox21 due to errors encountered in the public domain and in chemical supplier-provided information associated with chemical procurements. In addition, the accurate association of chemical structures to plated samples and assay results is a basic requirement of any cheminformatics or structure-activity relationship (SAR) modeling objectives associated with the ToxCast/Tox21 research programs. However, once the chemical contents in the original bottle has been suitably established, chemical analysis of neat and plated solutions provides an experimental standard of verification. Analytical QC is required to confirm the chemical identity and purity in the plated DMSO solutions undergoing testing at the time of plating, as well as at later time points (to assess sample stability over time).

**2.5.1 Analytical QC:** High-throughput LC-MS is the standard industry approach to analyzing HTS microtiter plates containing small solution volumes (typically 20-100ul) of hundreds of compounds, such as employed in ToxCast and Tox21 testing. The approach is cost effective and efficient in meeting the objectives of an HTS testing program, and is capable of providing useful information for the majority of plated samples.

Analytical QC procedures to establish purity, identity, concentration and stability of all plated Tox21 samples, including the complete set of EPA Tox21 library containing ToxCast Phase I\_v2, Phase II and E1K, are being carried out in association with the Tox21 program under an NTP-funded, NCGC-administered contract with OpAns Analytical Laboratory, located in Durham, NC. A full set of 384 well Tox21 parent plates, identical to those undergoing Tox21 assay testing, were submitted at the start or assay testing, i.e. time zero (t=0), for high throughput LC-MS analysis. The concentration chosen for the analytical analysis was 3mM using a volume of 20ul. Those passing identity (parent MW) checks with purity greater than 50% (Grade A, >90%, Grade B, 75-90%, Grade C, 50-75%, etc.) are not subject to further t=0 analysis. Those failing the identity or purity check, or for which no usable results are generated and the LC-MS method is deemed unsuitable (such as for low MW compounds, metals, etc.), undergo follow-up GC-MS at the National Institutes of Standards & Technology (NIST). Other failed compounds are potentially subject to follow-up LC-MS testing to increase the effective MW range, improve detection of polar compounds, and confirm insoluble samples using Flow Injection Analysis (FIA). An initial review of the LC-MS chromatograms is carried out by OpAns, with follow-up review,

ordering of additional analysis, and final analytical QC Grade assignments provided by an NCGC analytical chemist experienced with HTS operations. The overall process is summarized in [Figure 7](#).



**Figure 7.** General analytical QC approach for analysis of Tox21 plates.

In addition to the initial set of analytical QC plates analyzed at  $t=0$ , a second identical set of Tox21 plates, stored at room temperature under the same conditions as those being screened in Tox21 assays, is analyzed at  $t=4$  months to assess sample stability over time across the entire Tox21 compound library. For the subset of samples passing identity/purity checks at  $t=0$  but failing at  $t=4$  months, follow-up testing may be carried out for  $t=3$  months to establish a useful lifetime. This information will be used to inform subsequent assay analysis, and to set an overall “expiration” date on the Tox21 plates undergoing assay testing. Finally, a summary report of the QC analytical results, accompanied by the final QC grade for each Tox21 ID solution-level sample, will be made publicly available to inform the use and interpretation of Tox21 assay data (see [Figure 8](#)). [Note that preliminary Tox21 high-level summary QC grades are provided with the recent Phase II data release.]

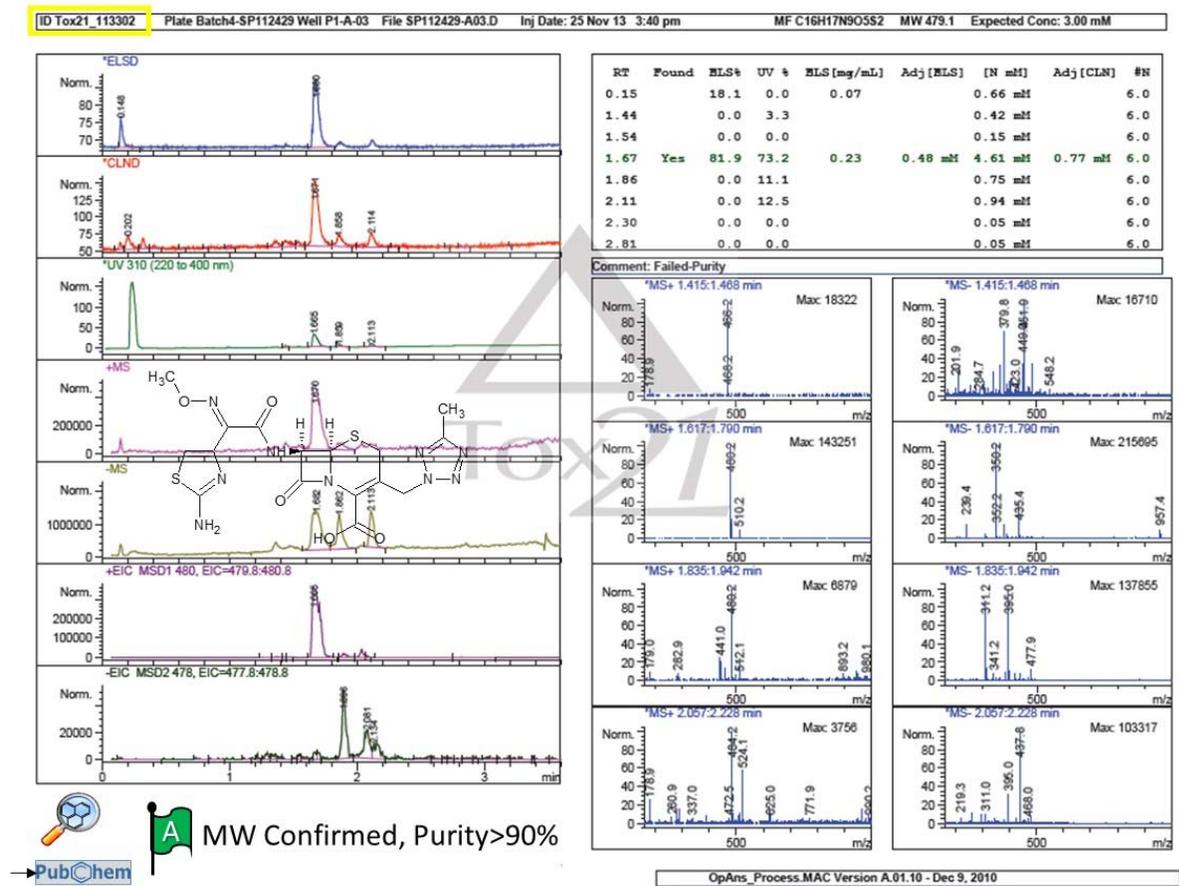


Figure 8. Mock-up pdf template for public release of Tox21 analytical QC results for each Tox21 ID sample, including the QC purity “Grade”, as well as an image of the chemical structure.

Final chromatograms and QC Grades have been completed for over 7K plated Tox21 samples from the original 10K Tox21 sample library (at the stock solution Tox21 ID level), with the remainder in the final stages of completion of GC-MS follow-up at NIST (approx. 1800) or undergoing customized method analysis at OpAns. Public release of the first batch of Tox21 summary pdfs (see Figure 8), along with a file containing the complete list of summary QC scores, is scheduled for early 2015 and will be accessible through PubChem, as well as the NIH Tox21 Chemical Browser (<http://tripod.nih.gov/tox21chem/>).

Figure 9 provides an early snapshot of the overall analytical QC results obtained for the 3 Tox21 partner sub-inventories (NCGC, NTP, EPA), illustrating the much higher proportion of “Inconclusives” associated with the substantially different chemical libraries, i.e., industrial and environmental chemicals vs. drugs. The plot in the lower right corner provides an indication that a large contributor to the Inconclusives

category for the EPA sub-inventory (and presumably for the NTP sub-inventory, as well) is the higher prevalence of low MW compounds vs. the NCGC drug library. Also, reassuring, is the very low rate of “Purity<50%” and “Fails” in the EPA Tox21 Inventory.

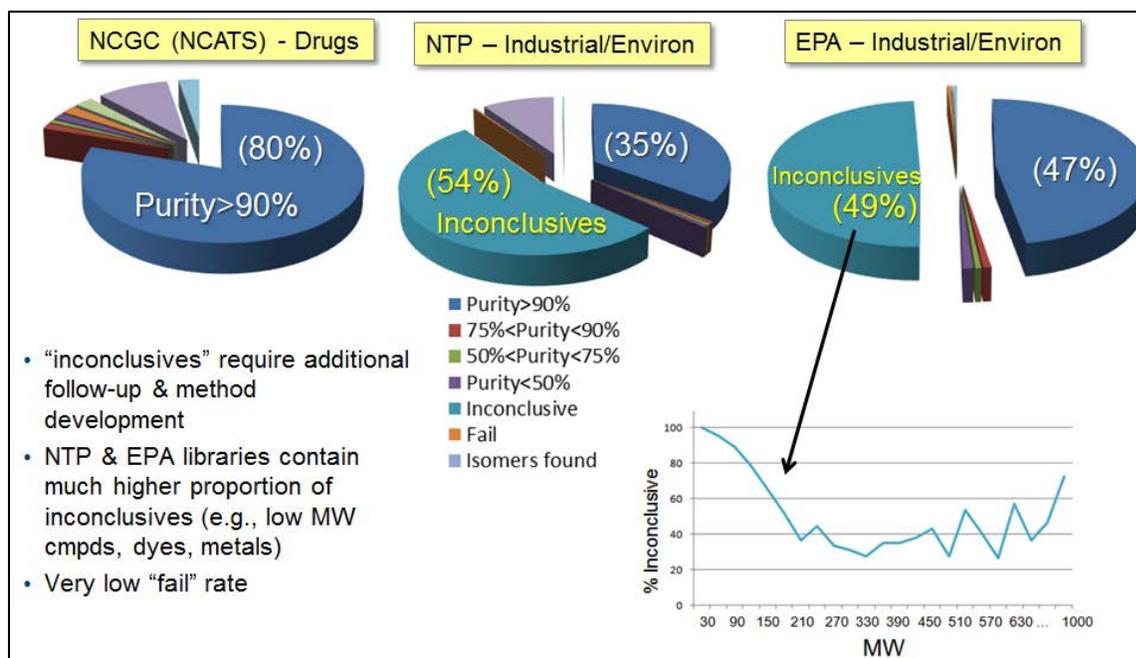


Figure 9. Snapshot of partial library Tox21 analysis results (completed as of 5/2014) comparing the results for the 3 Tox21 partner sub-inventories (NCGC, NTP, EPA).

**2.5.2 Tracking sample problems:** Solubility, or lack thereof, directly determines the effective concentration of compound delivered to a plate well and associated with an assay result. A sample can be deemed of high purity (Grade A) but be present at low concentration due to poor solubility or precipitation issues, thereby giving rise to false negative assay results due to low concentrations of chemicals. Another type of observation is that of a solution originally deemed “Soluble” and used for plating, and at a later time point reclassified as “Insoluble”, either due to precipitation or sample degradation over time.

Prior to 2013, EPA solubilizations were carried out by a single Contractor Operations (Ops) Project Leader, spanning creation of the entire Phase I Tox21 library (including Phase I, II, and E1K), with final DMSO solubility determined by visual inspection. This effectively enforced consistency in solubility determinations across the entire library. With retirement of this Ops Project Leader in 2013 and

replacement technicians tasked with performing EPA solubilizations, visual SOPs were introduced to provide greater consistency and clear guidance in determining solubility status under varied circumstances (e.g., hazy, clear supernatant with small amount of precipitate, etc). Accompanying these changes, EPA requested that additional solubility notes be added to the ComIT inventory report documenting relevant observations associated with marginal solubility calls (e.g., colored, hazy, cloudy, small amount of precipitate at bottom of vial, etc.). Such observations and calls are expected to be of use in triaging analytical QC runs and in isolating and flagging sample problems that could impact assay results.

A second source of potentially misleading assay results resulting from sample problems is compound volatility, which may lead to unexpectedly low concentration of chemical present in the plate well and false negative assay results. Volatility is a greater concern for low MW (and associated high vapor pressure and low boiling point) compounds, which are far more prevalent in the environmental/industrial chemicals included in EPA's (and NTP's) ToxCast and Tox21 inventories versus drugs in the NCGC Tox21 inventory. The impact of volatility can be difficult to assess in a testing environment as some low MW compounds (e.g., Formaldehyde) can be relatively stable in DMSO if immediately solubilized. Instances of stored neat samples later determined to be "empty on reweigh", have provided EPA with a more definitive means to label such potentially problematic chemicals, which may be deemed unsuitable for HTS testing.

Finally, there are the relatively rare reports from ToxCast vendors of observed sample problems detected during the course of testing. An example is a compound, whose structure and calculated phys-chem properties would be consistent with volatility concerns, suspected of having contaminated surrounding wells and producing confounding assay results. In another case, visible coloration and precipitate was observed within a plate well by an assay vendor.

In each of the above instances, EPA records the observational data in ChemInventory DB for use in informing future selection of chemicals for procurement or testing, as well as for annotating chemical-assay results during assay analysis. In addition, subsequent to procurement, upon further review of COA information, a small number of chemicals are deemed too volatile or reactive (e.g., evidence of stench) or otherwise unsuitable for testing. Again, such information is captured and recorded in ChemInventory DB for informing later procurements and testing selection.

Although ChemInventory DB is primarily intended as a chemical sample tracking database, it carries some compound-level information linked to the DSSTox\_GSID that flows from the sample handling and inventory designations. The distinction between sample and compound is an important one from the standpoint of chemical QC. Analytical QC results assessing sample identity and purity, for instance, are sample-specific results and are linked to the individual supplier, lot, age of sample, etc., which are also sample-specific attributes. ToxCast has attempted to minimize supplier-lot variation in samples through procurement of relatively large quantities of sample from primary suppliers (2 x 100mg), sufficient for spanning a complete phase of testing (e.g., Phase II). On the other hand, analytical QC indicating sample degradation over time, or attributes such as volatility, limited solubility, or reactivity, can be reasonably expected to be compound-level (i.e., DSSTox\_GSID) properties and, thus, be independent of supplier, lot, etc.

Within ChemInventory DB, a "GSIDxInventory" table carries compound-level information indicating the presence or absence (1 or 0) of a compound within any particular phase or sub-inventory of ToxCast (Phase I\_v1, Phase I\_v2, Phase II, Phase III, etc.) or Tox21 (including NTP or NCGC inventories). This table also carries historical notes pertaining to volatility, reactivity, limited solubility and evidence of degradation. Of the approximately 5K unique DSSTox\_GSIDs currently contained within ChemInventory DB (not counting non-EPA Tox21 Sample IDs), total numbers of compounds with identified problems, deemed unsuitable for future procurement or plating, include (as of 5/14/2014):

- 496 DMSO insolubles (down to 5mM)
- 119 volatiles (either empty on reweigh, or strong evidence for volatility)
- 34 compounds with stability problems, i.e. evidence for decomposition in DMSO over time
- 22 highly reactive

Retaining institutional "memory" of such information within ChemInventory DB, primarily gained during the course of sample handling and review, is considered a high priority to the overall chemical QC program. Such historical records flag potentially problematic compounds either for exclusion from testing (DO NOT PLATE) or for special attention, i.e. targeting analytical QC towards analysis of compounds of greatest concern, i.e. those of low solubility, low MW, or suspected to be susceptible to acid hydrolysis or degradation. In particular, such knowledge gained during the course of testing plated compounds is being captured and retained to inform the analysis of ToxCast and Tox21 HTS assay results further along the ToxCast data pipeline.

### 3. CONCLUSIONS – Chemical QC meeting practical and evolving needs

Chemical QC is a vital and continuing concern that is subject to varied practical constraints (cost, time, method limitations, etc.), and that has evolved over the course of EPA's ToxCast/Tox21 program to deal with the changing needs and demands of the program. With increasing success and broadened interest has come heightened pressures to screen more diverse chemicals, or particular chemicals or chemical lists of high environmental or toxicological concern, along with the greater uncertainties possibly associated with testing of such chemicals due to sub-optimal phys-chem properties. At the same time, particular assay results for individual chemicals of high regulatory or toxicological concern may be subject to greater interest and scrutiny, requiring a higher level of certainty associated with the assay result. These point to examples where greater reliance on chemical QC and a higher measure of assay validation may be required. This is both a hallmark and distinguishing factor of the ToxCast and Tox21 programs versus traditional pharmaceutical industry use of HTS assays for identifying a small number of candidate drugs (i.e., target hits) within a large chemical library for further study and testing. This also distinguishes the ToxCast and Tox21 programs from the large NIH MLP testing program, which is screening hundreds of thousands of chemicals to identify potentially bioactive compounds to support research to develop new therapeutics.

Requests of new ToxCast assay vendors or collaborators to receive ToxCast libraries at higher concentration DMSO solutions, to meet the needs of more complex cell-based assays for lower relative DMSO concentrations, also introduces greater potential uncertainties, since a higher percentage of compounds will be solubilized at or near their maximum achievable solubility in DMSO, putting these at potentially greater risk for precipitating during storage or plating. It is becoming increasingly important in such cases, as well as for volatile or marginally soluble chemicals, to institute analytical QC measures to assess effective concentration of a chemical in DMSO under plating conditions.

At the time of this writing, it has been more than 5 years since the conclusion of the pilot phase of ToxCast (Phase I\_v1), i.e. since the reprourement of the Phase I library as well as procurement of the EPA Tox21 library, from which Phase II chemicals were selected. Hence, the age of neat samples (powder or pure liquid) and the condition of stored/frozen solutions are of increasing concern. While attempts are being made to reprocore many chemicals in the original ToxCast library, measures are concurrently being instituted to better document and track the age and integrity of available solutions. Given the high expense and time investment involved in construction of EPA's ToxCast/Tox21 library, and the difficulty of locating sources for many chemicals of high interest to EPA, there is strong

motivation to prolong the lifetime of available samples to the extent possible and reasonable, while ensuring that degraded or questionable samples are identified and retired as efficiently as possible. Hence, new annotations are being added to ChemInventory DB to track both the age (age as indicated on the original COA, as well as reflecting the date of purchase), as well as the number of historical freeze/thaws of particular solutions (get ref from Keith), measures that would trigger analytical QC of the stock solution, or in the case of known decomposition problems (from Tox21 Phase 1 results), trigger disposal of the sample. The majority of problems are believed to be confined to aging DMSO solutions and the absorption of water over time for chemicals that are susceptible to acid hydrolysis and degradation, whereas the neat sample, stored frozen under inert conditions, can more reliably be used to create fresh stock solutions that retain the original properties of the newly procured chemical and restart the clock ( $t=0$ ). In addition, new solutions made from newly procured supplier/lot chemicals, which differ from those previously plated as part of EPA's Tox21 library, will be undergoing separate high-throughput LC-MS (minimum) to establish baseline ( $t=0$ ) analytical results for the current ToxCast chemical library.

Each of the above examples underscores that while the ultimate goal of chemical QC may be accurate and reliable information on each and every chemical undergoing ToxCast/Tox21 testing, there are practical limits as well as a balance that must be struck between cost/time/efficiency and the larger objectives of EPA's ToxCast/Tox21 program. In summary, in the absence of perfect knowledge and certainty, the ToxCast chemical QC process strives to minimize controllable sources of errors, particularly in the chemical information QC review and registration process, but also in handling and storage procedures. At the same time, the ToxCast and Tox21 analytical QC processes attempt to detect, understand, document and communicate actual errors and problems impacting particular chemicals under testing conditions. Armed with such knowledge, the larger objectives of EPA's ToxCast/Tox21 programs to screen a broad range of environmental chemicals for potential toxicity and to improve our understanding of the biological basis for such toxicity, has a greater probability of success.

## REFERENCES

Collins, F. S., Gray, G. M., Bucher, J. R. 2008. Toxicology. Transforming environmental health protection. *Science* 319: 906-907.

Dix, D.J., Houck, K.A., Martin, M.T., Richard, A.M., Setzer, R.W., Kavlock, R.J. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci.* 95, 5-12.

EPA EDSP21 Work Plan. 2011. Endocrine Disruptor Screening Program for the 21st Century: The Incorporation of In Silico Models and In Vitro High Throughput Assays in the Endocrine Disruptor Screening Program (EDSP) for Prioritization and Screening, U.S. EPA. Available at:

[http://www.epa.gov/endo/pubs/edsp21\\_work\\_plan\\_summary%20overview\\_final.pdf](http://www.epa.gov/endo/pubs/edsp21_work_plan_summary%20overview_final.pdf)

Judson, R., Richard, A., Dix, D., Houck, K., Elloumil, F., Martin, M., Cathey, T., Transue, T., Spencer, R., Wolf, M., 2008. ACToR - Aggregated Computational Toxicology Resource, *Toxicol. Appl. Pharmacol.* 233:7-13.

Judson, R., Richard, A., Dix, D.J., Houck, K., Martin, M., Kavlock, R.J., Dellarco, V., Henry, T., Holderman, T., Sayre, P., Tan, S., Carpenter, T., Smith, E. 2009. The Toxicity Data Landscape for Environmental Chemicals. *Environ. Health Perspect.* 117, 685-695.

Martin M.T., Judson R.S., Reif D.M., Kavlock R.J., Dix D.J. 2008. Profiling Chemicals Based on Chronic Toxicity Results from the U.S. EPA ToxRef Database. *Environ. Health Perspect.* 117:392-399.

Kavlock, R., Chandler, K., Houck, K., Hunter, S., Judson, R., Kleinstrauer, N., Knudsen, T., Martin, M., Padilla, S., Reif, D., Richard, A., Rotroff, D., Sipes, N., Dix, D. 2012. Update on EPA's ToxCast Program: Providing high throughput decision support tools for chemical risk management. *Chem. Res. Toxicol.* 25:1287-1302.

Knudsen, T.B., Houck, K.A., Sipes, N., Singh, A.V., Judson, R., Martin, M.T., Weissman, A., Kleinststeuer, N., Mortensen, H.M., Reif, D., Rabinowitz, R., Setzer, W., Richard, A.M., Dix, D.J., Kavlock, R.J. 2011. Activity profiles of 309 ToxCast™ chemicals evaluated across 292 biochemical targets. *Toxicol.* 282:1-15.

Richard, A.M. 2004. DSSTox Website launch: Improving public access to databases for building structure-toxicity prediction models. *Preclinica* 2:103-108.

Sipes, N.S., Martin, M.T., Kothiya, P., Reif, D.M., Judson, R.S., Richard, A.M., Houck, K.A., Dix, D.J., Kavlock, R.J., Knudsen, T.B. 2013. Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. *Chem. Res. Toxicol.* 26:878-895.

Tice, R.R., Austin, C.P., Kavlock, R.J., Bucher, J.R. 2013. Improving the human hazard characterization of chemicals: A Tox21 update. *Environ. Health Perspect.* 121, 756-765.