User's Guide for the Chemical Transformation Simulator (CTS), Version 2.0

04/03/2023

Chemical Transformation Simulator: A Cheminformatics Tool for Predicting Transformation Pathways and Physicochemical Properties

2023 U.S. Environmental Protection Agency

NOTICE

This document has been reviewed by the U.S. Environmental Protection Agency, Office of Research and Development, and approved for publication. It does not represent and should not be construed to represent any Agency determination or policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Purpose

The Chemical Transformation Simulator (CTS) User's Guide is designed to provide the first-time user a complete understanding of how to utilize the CTS tool. The User's Guide may be reviewed from start to finish or by moving directly to a topic of interest through selection of the appropriate topic in the Table of Contents.

Table of Contents

1. Introduction

The Chemical Transformation Simulator (CTS) provides the calculated physicochemical properties of a target chemical and its transformation products, which are predicted as a function of the reaction system of interest. This is accomplished through the integration of cheminformatics applications for the encoding of process science underlying transformation pathways and computational chemistry tools for the calculation of physicochemical properties.

The CTS consists of three modules, the selection and order of execution of which is based on the user's choice of one of three available workflows as described below.

- **Chemical Editor (CE) Module:** Provides options for chemical entry through SMILES notation, chemical name, CAS #, or drawn structure, as well as speciation of the parent chemical
- **Physicochemical Properties (PCP) Module:** Calculates physicochemical properties for the parent chemical and predicted transformation products based on the executions of multiple physicochemical calculators
- **Reaction Pathway Simulator (RPS) Module:** Generates potential transformation products based on user-specified reaction conditions

2. Background

A key Agency need identified as a high priority in the Chemical Safety for Sustainability (CSS) National Research Program is for high throughput computational systems to rank chemicals based on hazards and risks. To fully characterize the risks associated with the release of a chemical into the environment, tools are necessary to simulate environmental fate and transport for organic chemicals for which such data are not available. Knowledge of inherent chemical properties (ICP) is essential for the parameterization of environmental fate and transport models. Of the ~85,000 chemicals in the TSCA inventory, it is estimated that high quality measured ICP data are available for less than 2% of these chemicals. Additionally, 20 to 30 new chemicals a month are being assessed through the Office of Pollution Prevention and Toxics (OPPT) Pre-Manufacturing Notification (PMN) process. This ever-growing data gap must be addressed through the development of a high throughput computational system for calculating the ICP necessary for the parameterization of environmental fate models used to estimate environmental

concentrations of both the parent chemical and predicted transformation products as a function of environmental conditions.

The key components of the CTS are the Physicochemical Properties module (PCP) and the Reaction Pathway Simulator module (RPS). The PCP is based on a consensus approach that allows the user to compare output generated by several calculators that take different approaches to calculating specific physicochemical properties. These calculators include (1) EPI Suite™, which uses a fragment-based approach, (2) T.E.S.T. (Toxicity Estimation Software Tool), which uses Quantitative Structure Activity Relationship (QSAR)-based approaches, (3) ChemAxon plug-in calculators, which use an atom-based fragment approach, and (4) OPERA (OPEn structure-activity/property Relationship App), which uses a weighted k-nearest neighbor approach to construct QSAR and Quantitative Structure Property Relationship (QSPR) models. The output derived from these calculators will enable the user to compare the calculated data with measured data in readily accessible web-based databases.

The output of the RPS is based on the selection and execution of reaction libraries that represent one-step reactions for transformation (e.g., abiotic reduction and hydrolysis) of reactive functional groups. These one-step reactions represent viable transformation pathways based on the identification and subsequent transformation of reactive functional groups. A reaction library for human phase I metabolism developed by ChemAxon is also available through the CTS. The development of reaction libraries allows us to "encode" the known process science published (current and future) in the peer-reviewed literature. The encoding of process science is accomplished using Chemical Terms Language and Smart Reaction strings through cheminformatics applications. The execution of these reaction libraries provides dominant transformation pathways and products for the chemical of interest as a function of environmental conditions.

3. Using the CTS Software

3.1. Restrictions

The CTS is designed to predict transformation pathways and calculate physicochemical properties for organic chemicals. The CTS is designed for users with expertise regarding the environmental processes controlling the degradation of organic chemicals and the use of physicochemical properties in environmental fate modeling. Currently, organometallics, non-dissociating salts of organic chemicals, and polymers are not recognized by the CTS. To provide CTS users with a rich user experience we take advantage of features specified in the HTML 5 standards. Internet Explorer does not fully implement support for HTML 5 standards. As a result, CTS functionality will be limited when using the Internet Explorer web browser. We recommend using the latest version of Google Chrome, Microsoft Edge, or Mozilla Firefox.

3.2. Accessing the CTS

CTS available to the public at https://qed.epa.gov/cts/. From the home page, the user can select one of three available CTS workflows, or access general information concerning the major modules of the CTS, the physicochemical calculators, and the reaction libraries, as shown in the screen shot below.

4. CTS Modules Overview

4.1. Chemical Editor (CE):

CTS's Chemical Editor (CE) appears at the beginning of all workflows and allows users to enter chemicals by their name, Chemical Abstracts Service Registry Number (CAS#), or Simplified Molecular-Input Line-Entry System (SMILES) string. Alternatively, the Ketcher open-source chemical structure editor is implemented in CTS to allow users to draw a chemical structure (EPAM Life Sciences; https://lifescience.opensource.epam.com/ketcher/). ChemAxon's JChem application is used to generate a standardized SMILES string, and EPA's CompTox Chemistry Dashboard (https://comptox.epa.gov/dashboard) is used to obtain the preferred common name, IUPAC name, chemical formula, relevant CAS numbers, average and monoisotopic masses, and the DTXSID (unique substance identifier assigned by EPA's Center for Computational Toxicology and Exposure (CCTE)) for the selected chemical.

4.2. Chemical Speciation:

CTS's Chemical Speciation (CS) workflow uses ChemAxon's Plugin Calculators to generate:

- The speciation of a chemical as a function of pH
- The ionization constant(s)
- The dominant tautomer distribution; and
- Structures for all possible isomers

4.3. Physicochemical Properties (PCP) Module:

CTS's Physicochemical Properties Module calculates physicochemical properties for the parent chemical and predicted transformation products based on the findings of multiple physicochemical calculators. The PCP is based on a consensus approach that allows users to compare output from multiple calculators that use different approaches to calculate specific physicochemical properties. Additionally, the geometric mean of the predicted values (not including any measured data or ionization constants) for each selected property from the selected calculators will be automatically calculated and displayed in the table when predicted values are requested.

The calculators that PPC is currently accessing include:

1. EPI Suite™, which uses a QSAR approach with either fragment counts or properties as descriptors;
2. Toxicity Estimation Software Tool (T.E.S.T.), which implements several approaches, including fragment-based QSARs, hierarchical clustering with structural similarity based on a variety of 2-D physicochemical descriptors, and nearest neighbor;

3. ChemAxon plug-in calculators, which are based on QSARs with either atom or fragment counts as descriptors; and
4. OPEn structure activity/property Relationship App (OPERA), which uses a weighted k-nearest neighbor approach to construct QSAR/QSPR models based on 2-D physicochemical descriptors.

Users also have the option to request measured data that is available in the EPI Suite™ PHYSPROP physicochemical property database.

4.4. Reaction Pathway Simulator (RPS):

CTS's Reaction Pathway Simulator (RPS) generates potential transformation products based on user-specified reaction conditions. The output of the RPS is based on the selection and execution of reaction libraries that represent reaction schemes for the transformation of reactive functional groups that are susceptible to processes such as reduction and hydrolysis. These reaction schemes denote viable transformation pathways based on the identification and transformation of the reactive functional groups. A rank is assigned to each one of the reaction schemes based on available experimental data. The rank is essentially a relative reaction rate, defined on a scale of one to seven, with seven being assigned to the fastest reaction schemes. The rank of each scheme is used to calculate an approximate percentage production of each potential transformation product.

In Metabolizer, an algorithm has been implemented to approximate a percent production and accumulation for products of each reaction scheme. In the Metabolizer algorithm, the unitless "formation" value for scheme i (f_i) is defined as the number 7 raised to the power of the rank:

$$f_i \stackrel{\text{def}}{=} 7^{Rank_i}$$

The "formation" values are analogous to rate constants, and the "production" (in %) of the product(s) generated by scheme i is calculated according to the following equation:

$$\% P_i = \frac{100 f_i}{\sum_{j=1}^{N} f_j}$$

where N is the total number of transformation schemes in the current generation. The "accumulation" of the product formed by scheme i is then calculated as the difference between f_i and the summation of the formation values for all M schemes that may transform the product in the next generation, normalized by the summation of the formation values for all transformation schemes in the current generation:

$$\% A_i = \frac{100\left(f_i - \sum_{l=1}^{M} f_l \right)}{\sum_{j=1}^{N} f_j}$$

A more detailed description of the production and accumulation calculations is provided in the online documentation for the CTS Reaction Libraries.

Developing reaction libraries allows scientists to "encode" the known process science published – current and future – in the peer-reviewed literature. Encoding process science is accomplished by using Chemical Terms Language and cheminformatics applications. Reaction libraries have been developed for the environmental transformation processes of abiotic hydrolysis and abiotic reduction and direct photolysis (unranked and ranked). A Spontaneous reaction library, which is intended to be combined with other CTS reaction libraries, is also available to capture the rapid transformation of intermediates that form due to other transformation processes (e.g., hydrolysis or photolysis). A reaction library is currently under development to predict the microbial transformation products of organic chemicals in anaerobic environments.

Two additional libraries have been developed to predict the likely environmental and metabolic transformation products of per- and polyfluoroalkyl substances (PFAS). These libraries are designed to run as standalone libraries and cannot be combined with other CTS reaction libraries.

In addition to the libraries developed for CTS, the RPS links to several externally developed transformation prediction tools. A reaction library for human phase I metabolism that was developed by ChemAxon is made available through the RPS. CTS provides likely aerobic biotransformation products from the EnviPath prediction system for the microbial biotransformation of organic environmental contaminants through web services. Finally, CTS accesses predictions from four independent BioTransformer modules: the Phase I (CYP450) transformer for predicting metabolism in mammals, the Phase II (CYP450) transformer for predicting conjugation reactions in mammals, the Gut Microbial Transformer for predicting gut microbial metabolism, and the EC-based Transformer for predicting promiscuous metabolism.

Executing these reaction libraries provides dominant transformation pathways and products for the chemical of interest as a function of environmental conditions. Users also have the option to execute the PCP for the calculation of physicochemical properties for the parent chemical and transformation products.

Documentation on the process science supporting the development of the CTS reaction libraries is available on the Reaction Libraries page by clicking on the appropriate link in the left-hand frame. For example, the documentation for the “Abiotic Hydrolysis Reaction Library” lists the 24 reaction schemes in the library, as shown in the screen shot below.

Abiotic Hydrolysis Reaction Library

The Abiotic Hydrolysis Reaction Library has been developed as a component of the Chemical Transformation Simulator (CTS), a web-based software tool under development in EPA's Office of Research and Development. The library is implemented in CTS to predict the likely hydrolytic transformation products in the environment for an organic chemical of interest.

Version 1.8 of the Abiotic Hydrolysis Reaction Library contains 24 reaction schemes:

- Halogenated Aliphatics: Nucleophilic Substitution
  - Scheme A: C-X with no other adjacent halogens
  - Scheme B: C-X with vicinal halogen atoms
  - Scheme C: C-X with geminal halogen atoms
- Halogenated Aliphatics: Elimination
- Epoxide Hydrolysis
- Organophosphorus Ester Hydrolysis 1 (Base-Catalyzed)
- Organophosphorus Ester Hydrolysis 2 (Neutral or Acid-Catalyzed)
- Carboxylic Acid Ester Hydrolysis
- Lactone Hydrolysis
- Carbonate Hydrolysis
- Cyclic Carbonate Hydrolysis
- Anhydride Hydrolysis
- Cyclic Anhydride Hydrolysis
- Amide Hydrolysis
- Lactam Hydrolysis
- Carbamate Hydrolysis
- Thiocarbamate Hydrolysis
- Urea Hydrolysis
- Cyclic Urea Hydrolysis
- Sulfonylurea Hydrolysis
- Nitrile Hydrolysis
- N-S Cleavage
- Imide Hydrolysis
- Acid Halide Hydrolysis

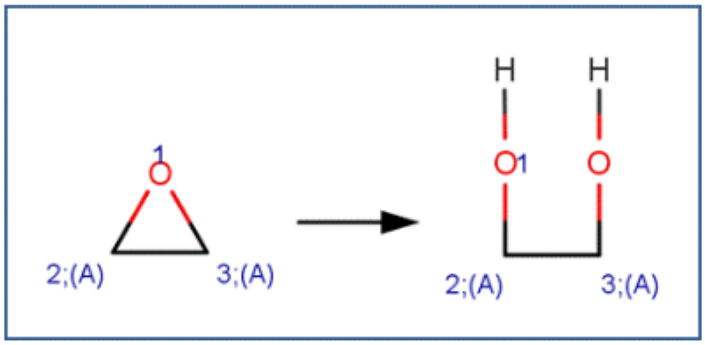
Ranking of Hydrolysis Reaction Schemes

Version History

Selection of one of the transformation pathways provides the reaction scheme, and documented examples with references. The following illustrates this information for Epoxide Hydrolysis:

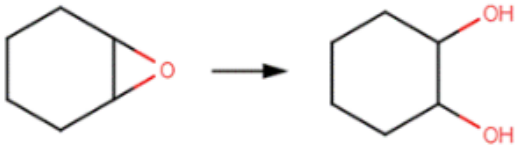
Epoxide Hydrolysis

SCHEME:



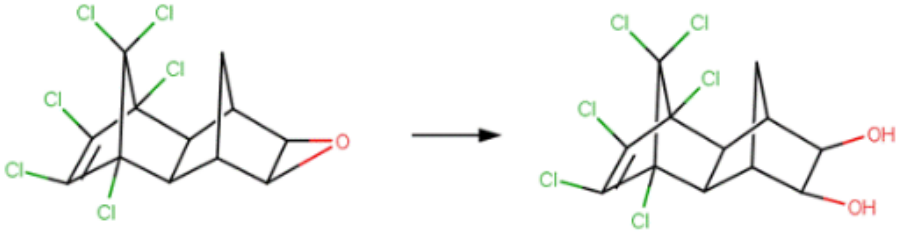
EXAMPLES:

- 1,2-Epoxycyclohexane (McMurry, 2011)

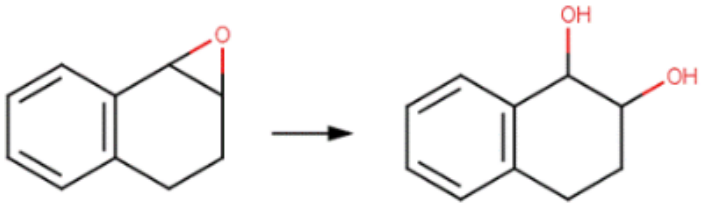


- Epichlorohydrin (Gaca *et al*, 2011)

- Endrin (Larson and Weber, 1994; U.S. EPA, 1992)



- 1,2-Epoxy-1,2,3,4-tetrahydronaphthalene (Becker *et al*, 1979)

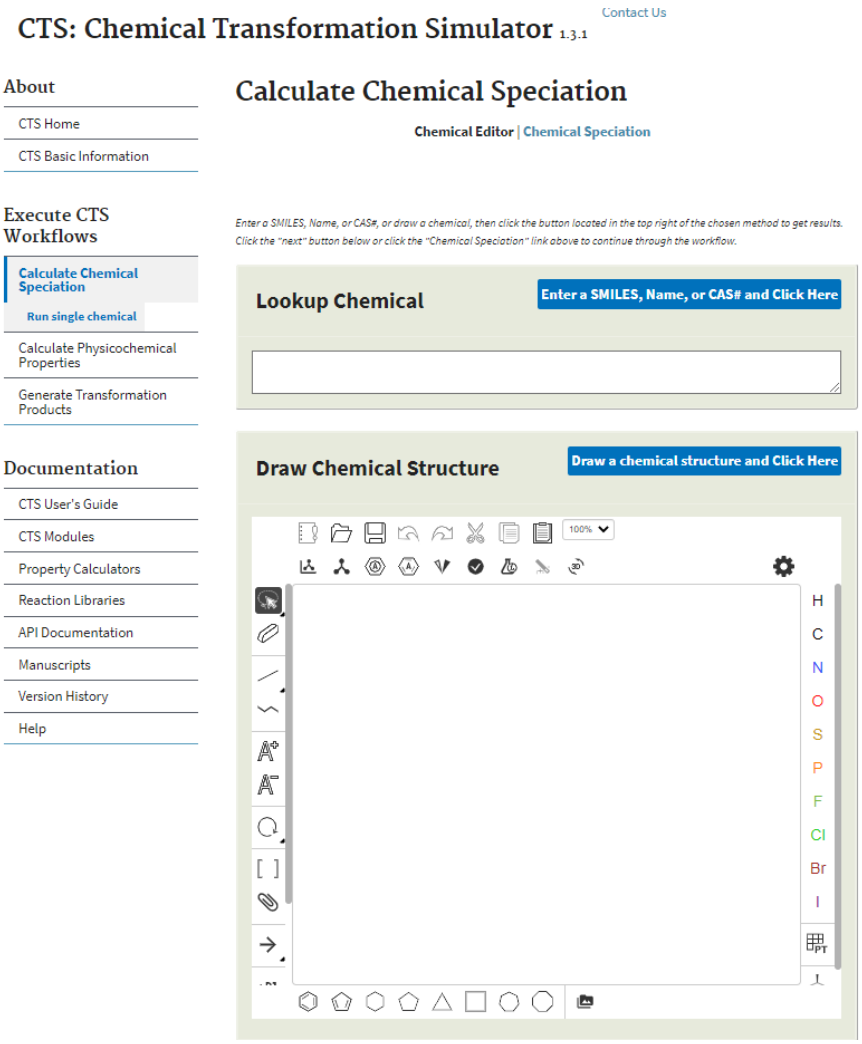


5. Execution of the CTS

The CTS is executed by selecting one of three available workflows (see descriptions below) and then entering a single chemical or a batch input file. The process for entering a single chemical or a batch input file, as described below, is identical for each of the workflows.

5.1. Single Chemical Entry

For single chemical entry, the "Run single chemical" tab is selected below the workflow of interest in the left-hand frame. The Chemical Editor appears where there is the option to either enter a SMILES String, name, or Chemistry Abstracts Service (CAS) # in the Lookup Chemical box, or to draw a chemical structure using the Chemical Editor (see screen shot below). For either case, the appropriate box must be clicked after providing the required information.

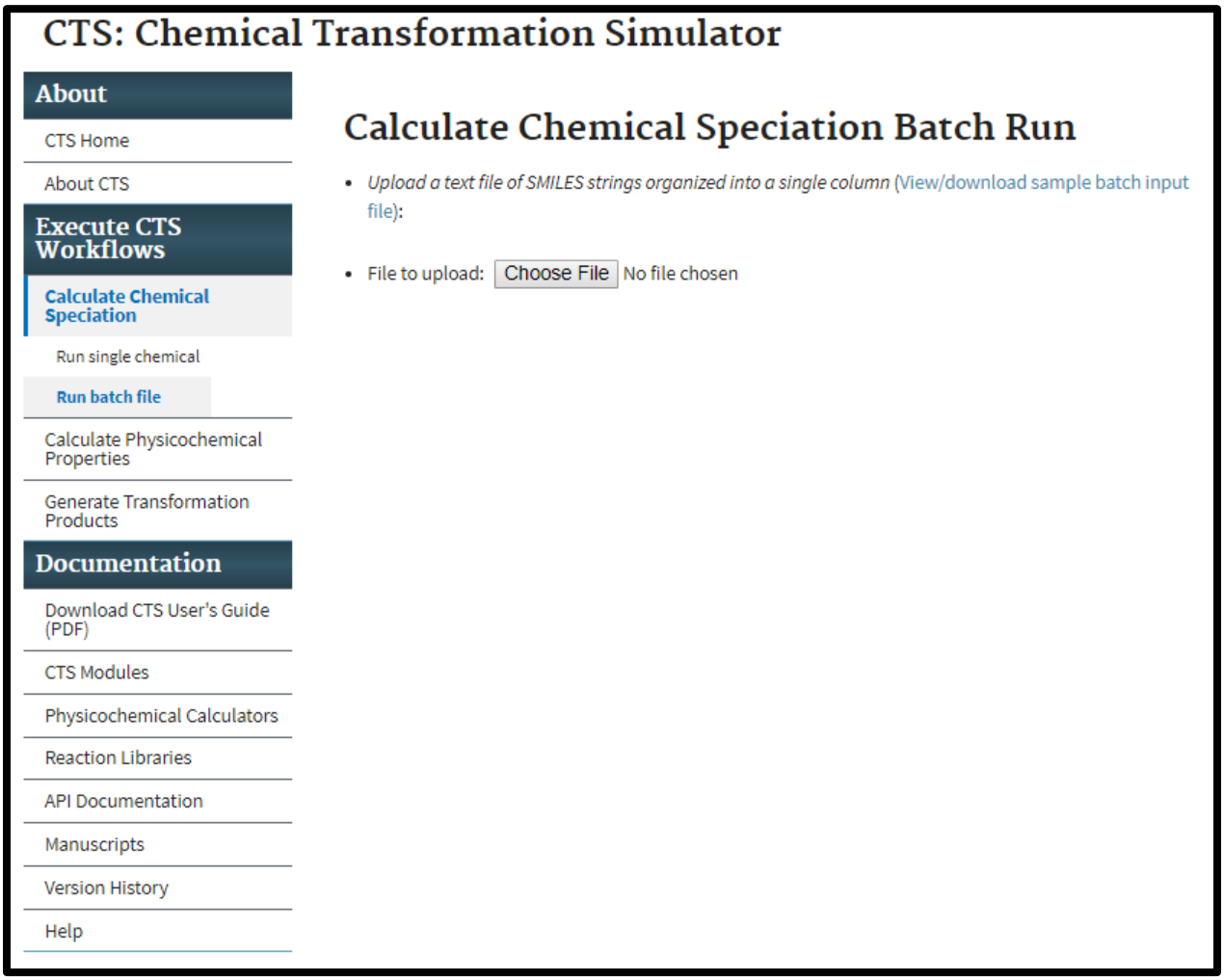


5.2. Batch Chemical Entry

For batch chemical entry, the "Run batch file" tab is selected below the workflow of interest in the left-side frame. By clicking on the sample batch input link, the example batch file, shown below in the screen shot, is opened or downloaded. The chemicals are entered as a single column of SMILES strings. The default value is currently set to a maximum of 10 chemicals.

Sample batch input:

```
C1=CC=CC=C1
CC(=O)OC1=CC=CC=C1C(O)=O
OC1=CC=CC=C1
OC(=O)CC(O)(CC(O)=O)C(O)=O
[O-][N+](=O)C1=CC=C(C=C1)[N+]([O-])=O
```



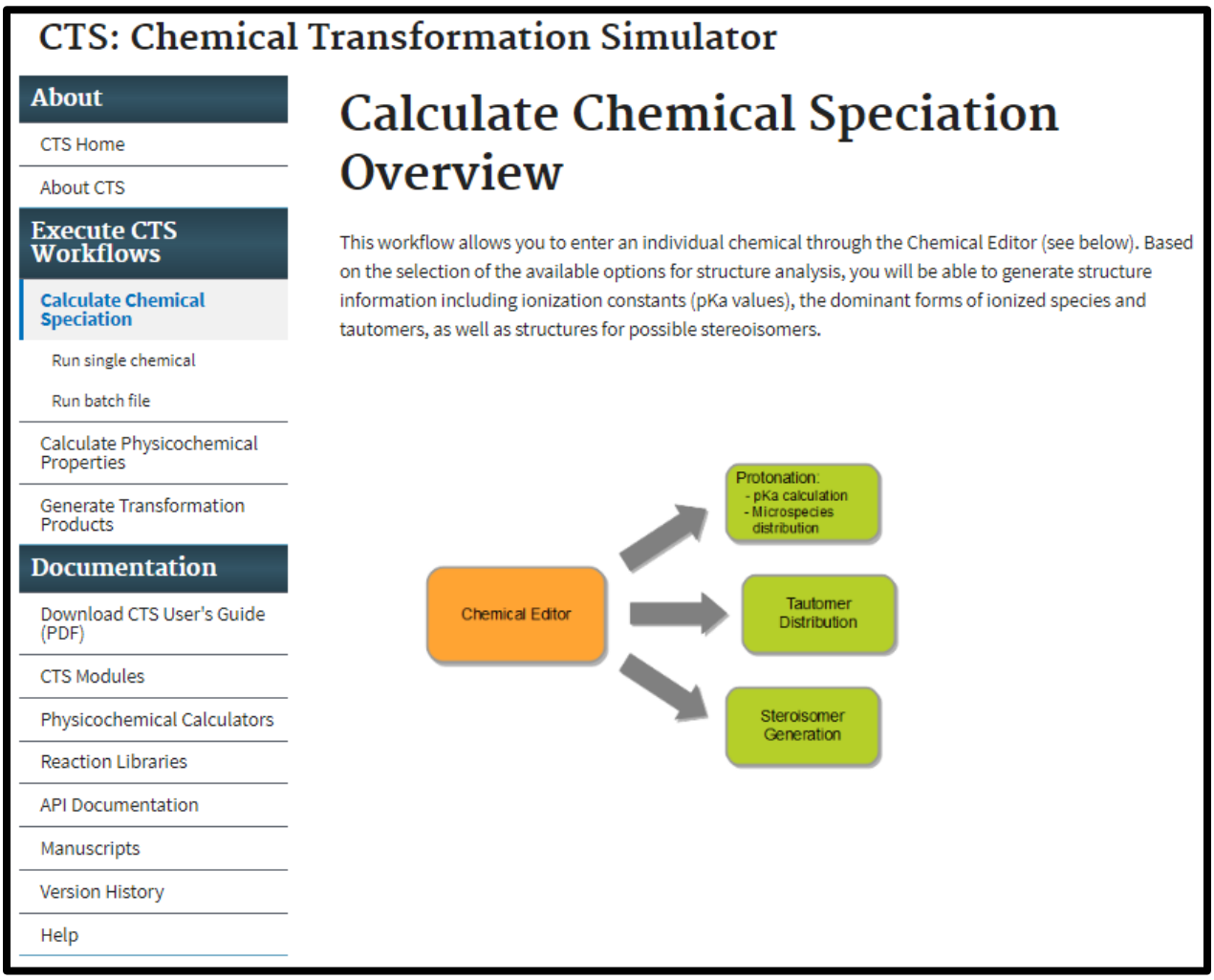
6. Execution of the CTS Workflows

The user executes the CTS through the selection of one of three available workflows:

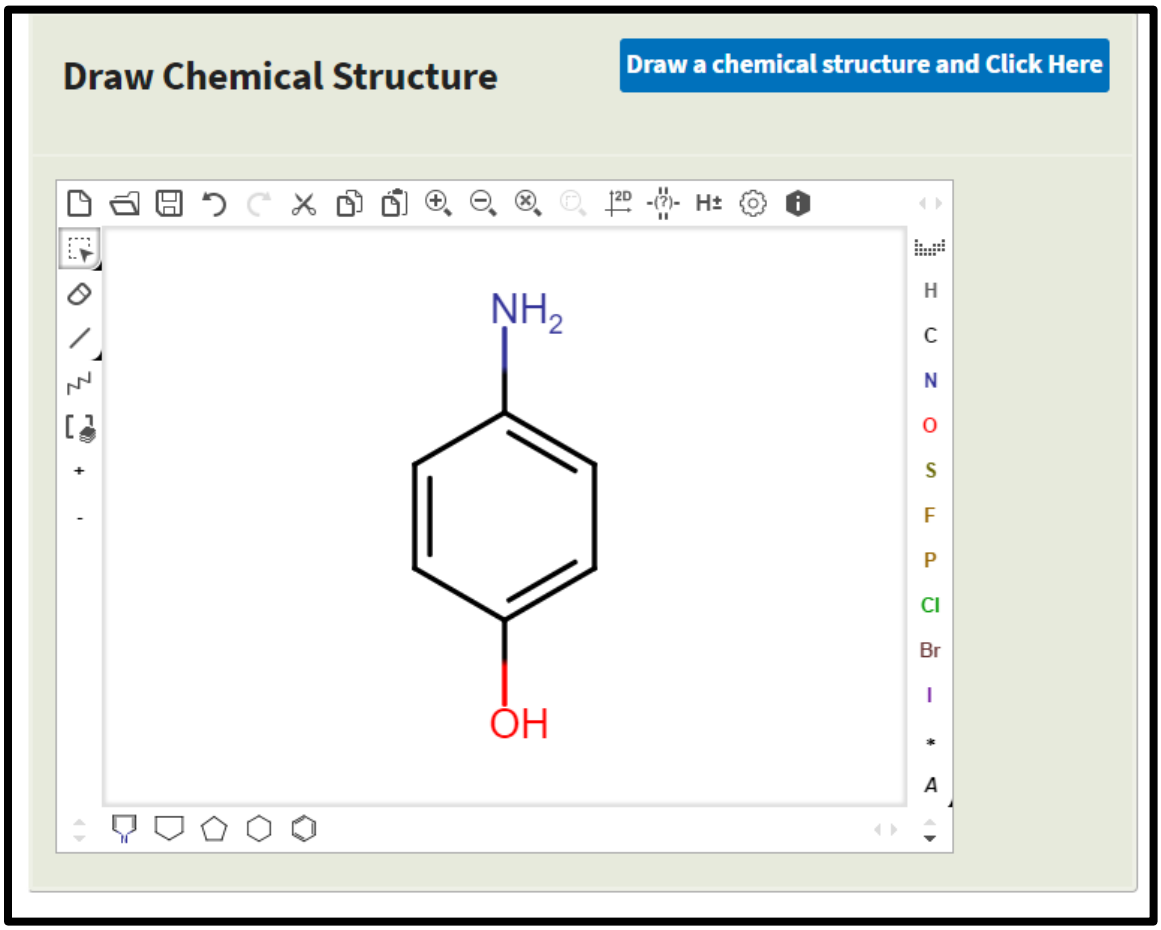
- Calculate Chemical Speciation
- Calculate Physicochemical Properties
- Generate Transformation Products

6.1. Calculate Chemical Speciation Workflow

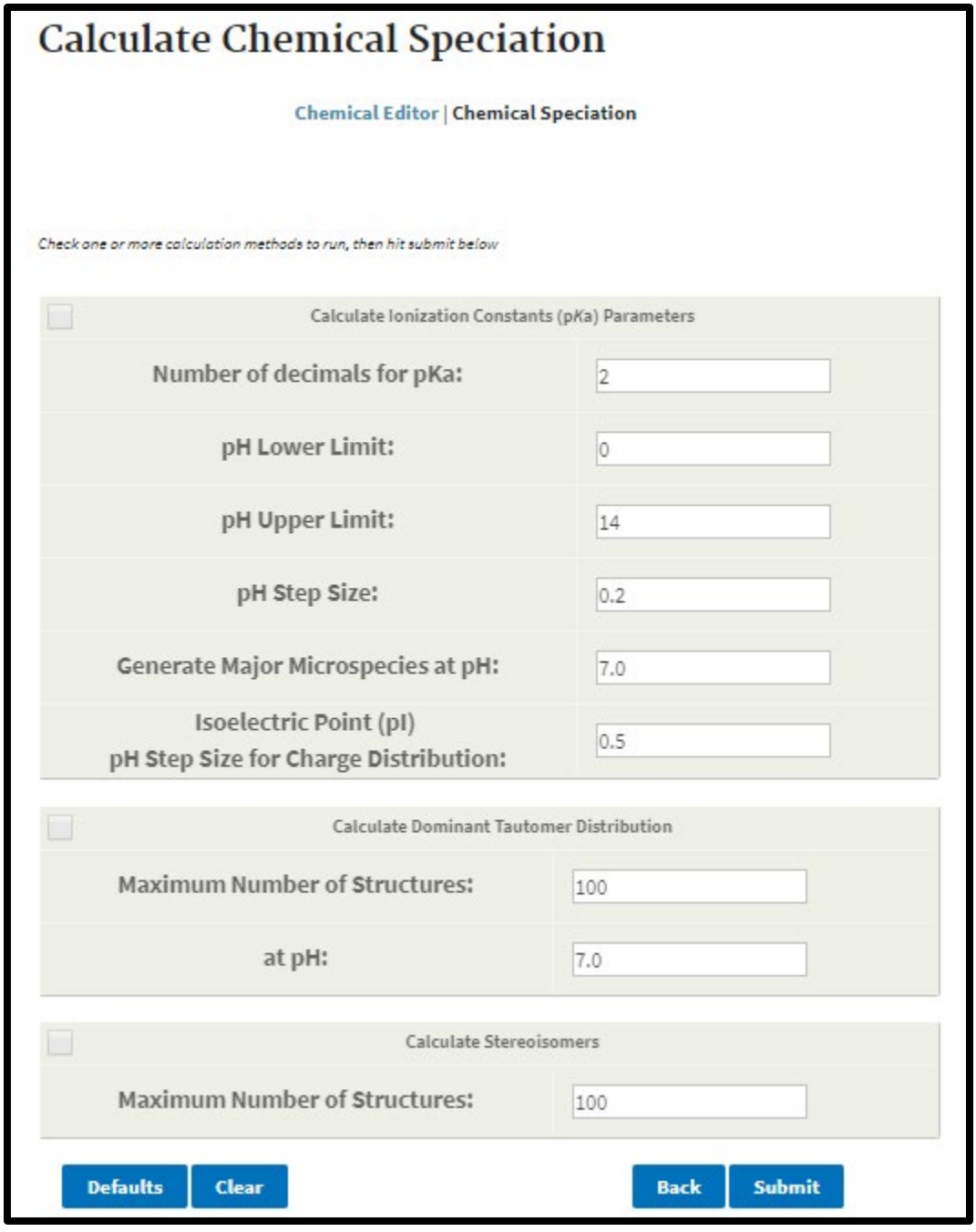
Selection of the Calculate Chemical Speciation Workflow provides this page illustrating the workflow overview as illustrated below. ChemAxon calculator plugins are executed for the calculation of chemical speciation.



Clicking on the “Run single chemical” link takes the user to the Chemical Editor. For the following example, 4-aminophenol was entered into the Chemical Editor as shown in the screen shot below.



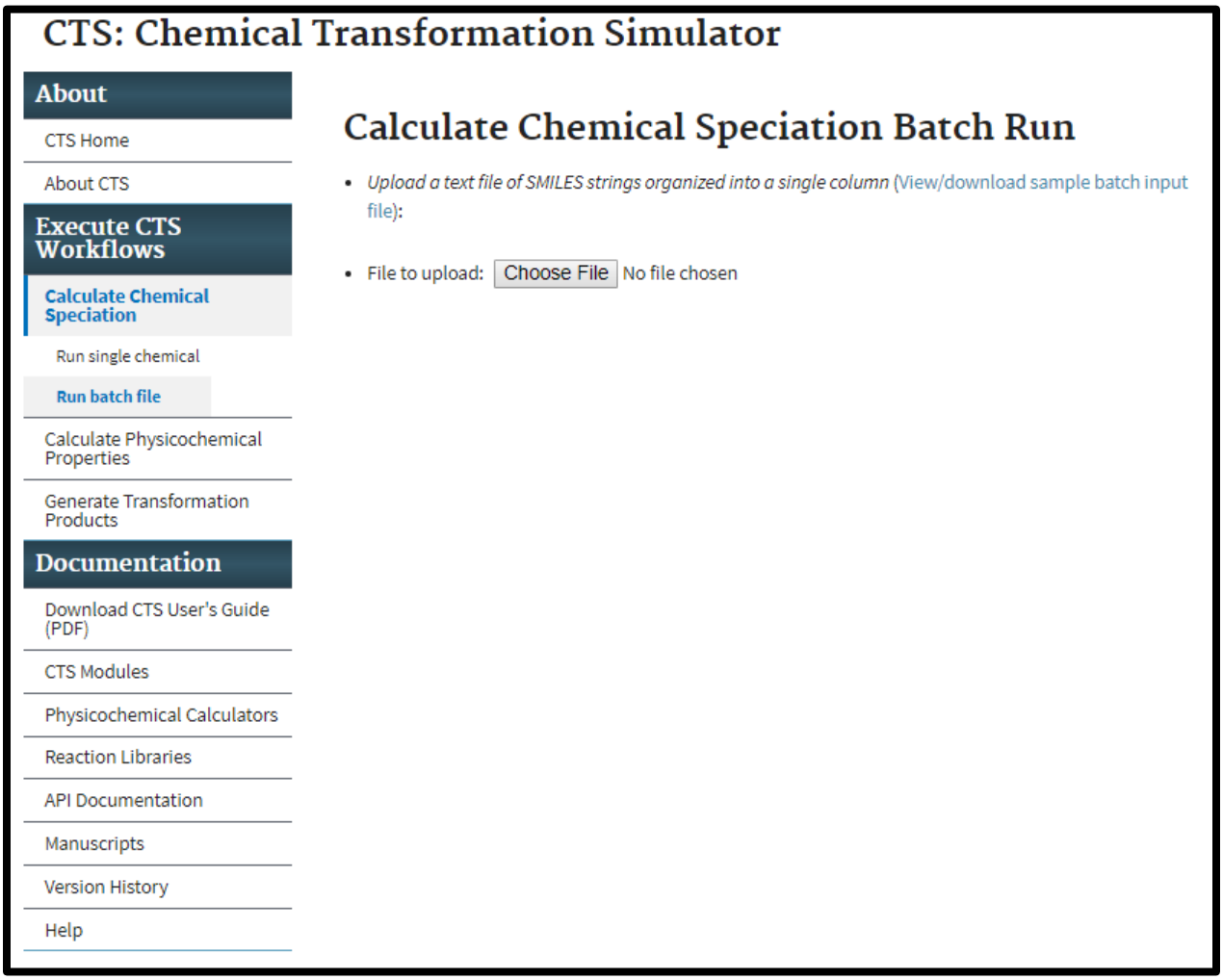
After clicking the Next button at the bottom of the Chemical Editor or the Chemical Speciation link at the top of the workflow frame, select from three available options for calculating chemical speciation as shown in the screen shot below.



Select any combination of the calculators; use the provided default values or change the default values required by the user. The following parameters can be adjusted:

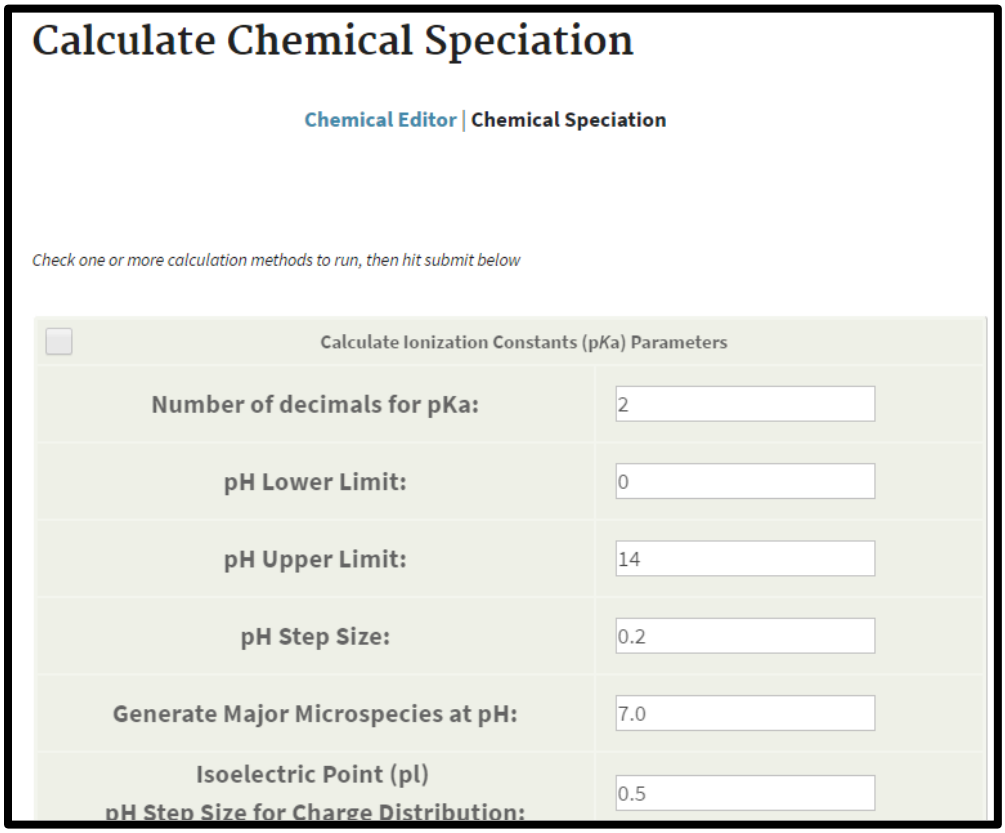
- **Calculate Ionization Constants**
  - Number of decimals: Number of decimal places calculated for acidic and basic pK_a values
  - pH Lower limit: Specifies the lower end of the pH range for which the microspecies will be generated
  - pH Upper limit: Specifies the upper end of the pH range for which the microspecies will be generated
  - Generate Major Microspecies at pH: Generates the Major Microspecies at the specified pH.
  - pH step size: Specifies the pH step size for the X-Axis of the plot illustrating the distribution of the microspecies as a function of pH
  - Isoelectric Point (pI) pH Step Size for Charge Distribution: Specifies the pH step size for the X-Axis of the plot illustrating the Isoelectric Point and charge distribution as a function of pH
- **Calculate Dominant Tautomer Distribution**
  - Maximum Number of Structures: Specifies the maximum number of structures that will be generated.
  - At pH: Specifies the pH at which the dominant tautomer distribution will be calculated
- **Calculate Stereoisomers**
  - Maximum Number of Structures: Specifies the maximum number of structures that will be generated.

The user also has the option of running a batch file. By clicking on the "Run batch file link", the following screen appears. The user has the option to view or download a sample batch input file or enter a file by clicking on the Choose File button.



6.2. Calculate Ionization Constants

Once the calculator(s) has been chosen and the appropriate parameters entered, click the Submit button to view the results. The calculator for ionization constants has been chosen for this demonstration as shown in the screen shot below.



The results of the ionization constant calculation are illustrated in the screen shots below:

- **pKa Calculations:** Provides the chemical structure entered, the generated microspecies, and the distribution of microspecies as a function of pH over the pH range specified. Results are color coded.
- **Isoelectric Point:** The isoelectric point is provided as well as a graph illustrating the charge on the chemical as a function of pH.
- **Major Microspecies:** The dominant microspecies formed at the pH selected.

- Results

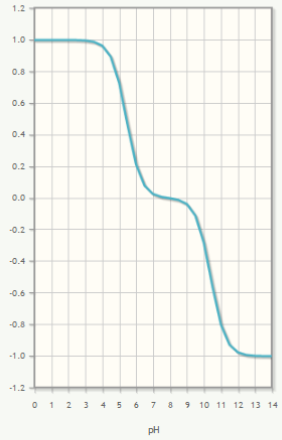
- pKa

Basic pKa Value(s): [5.43]

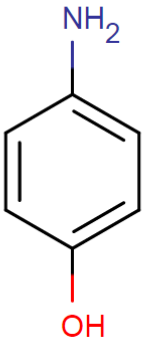
Acidic pKa Value(s): [10.4]

- **Isoelectric Point**

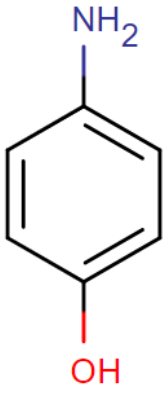
Isoelectric Point = 7.92



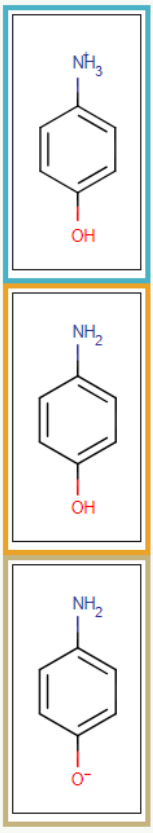
- **Major Microspecies at pH: 7.0**



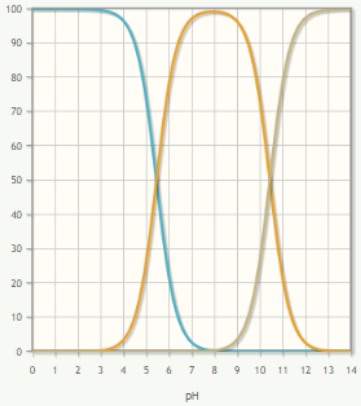
Parent



Microspecies

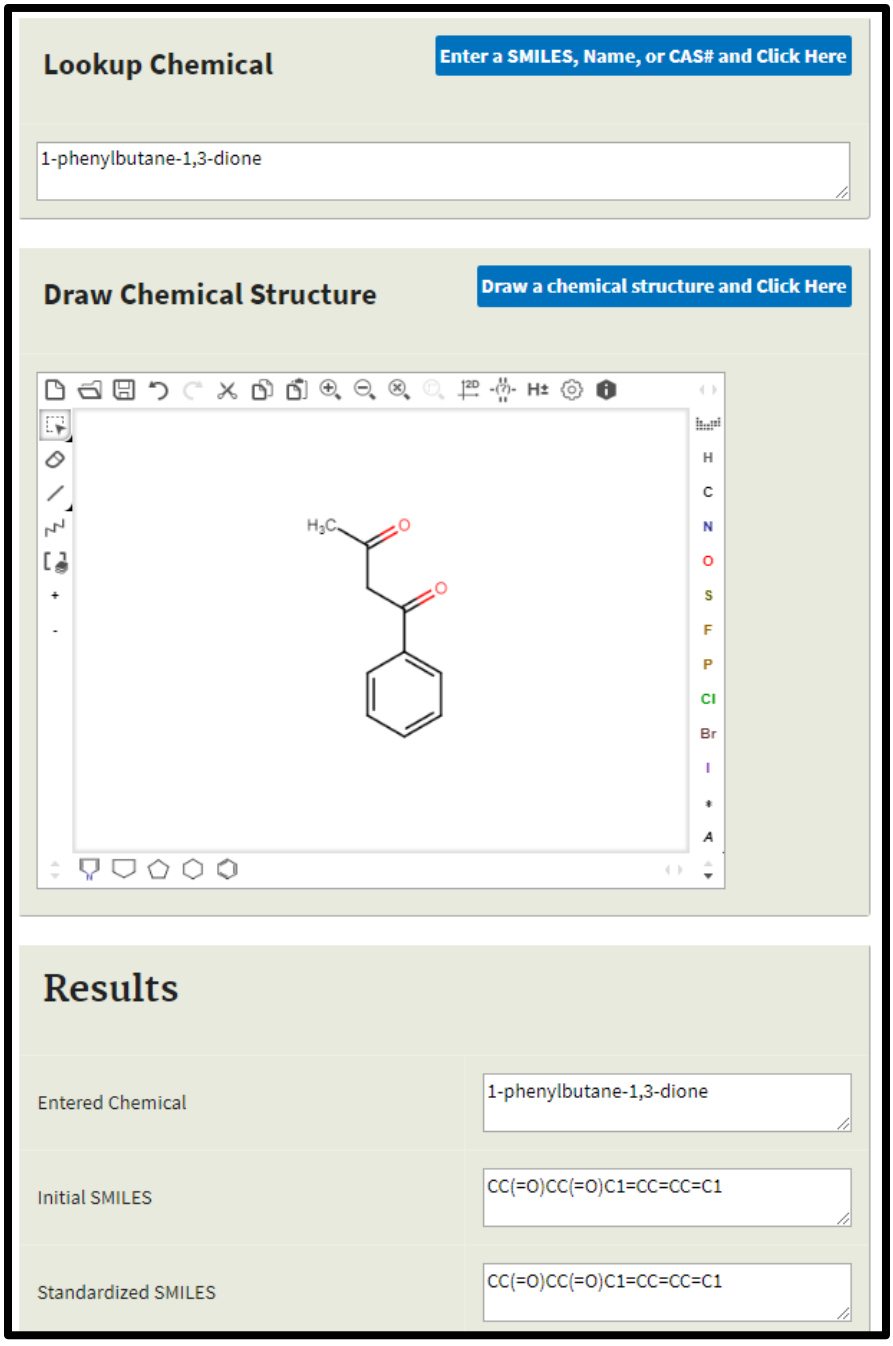


Microspecies Distribution (%)

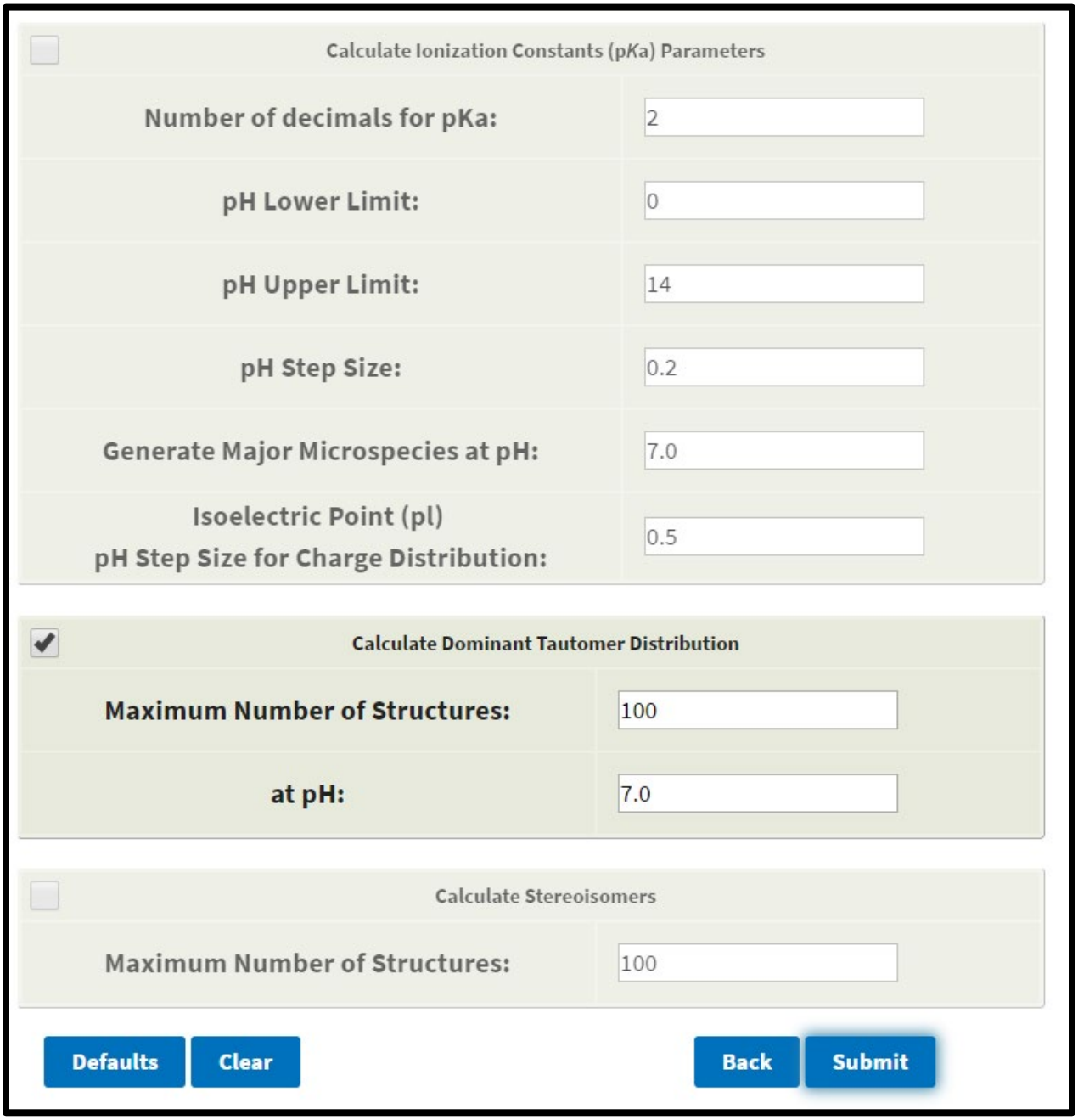


6.3. Calculate Dominant Tautomer Distribution

The CTS uses ChemAxon's tautomerization engine for the calculation of the dominant tautomer distribution. For this demonstration, 1-phenylbutane-1,3-dione has been entered into the Chemical Editor as shown in the screen shot below.

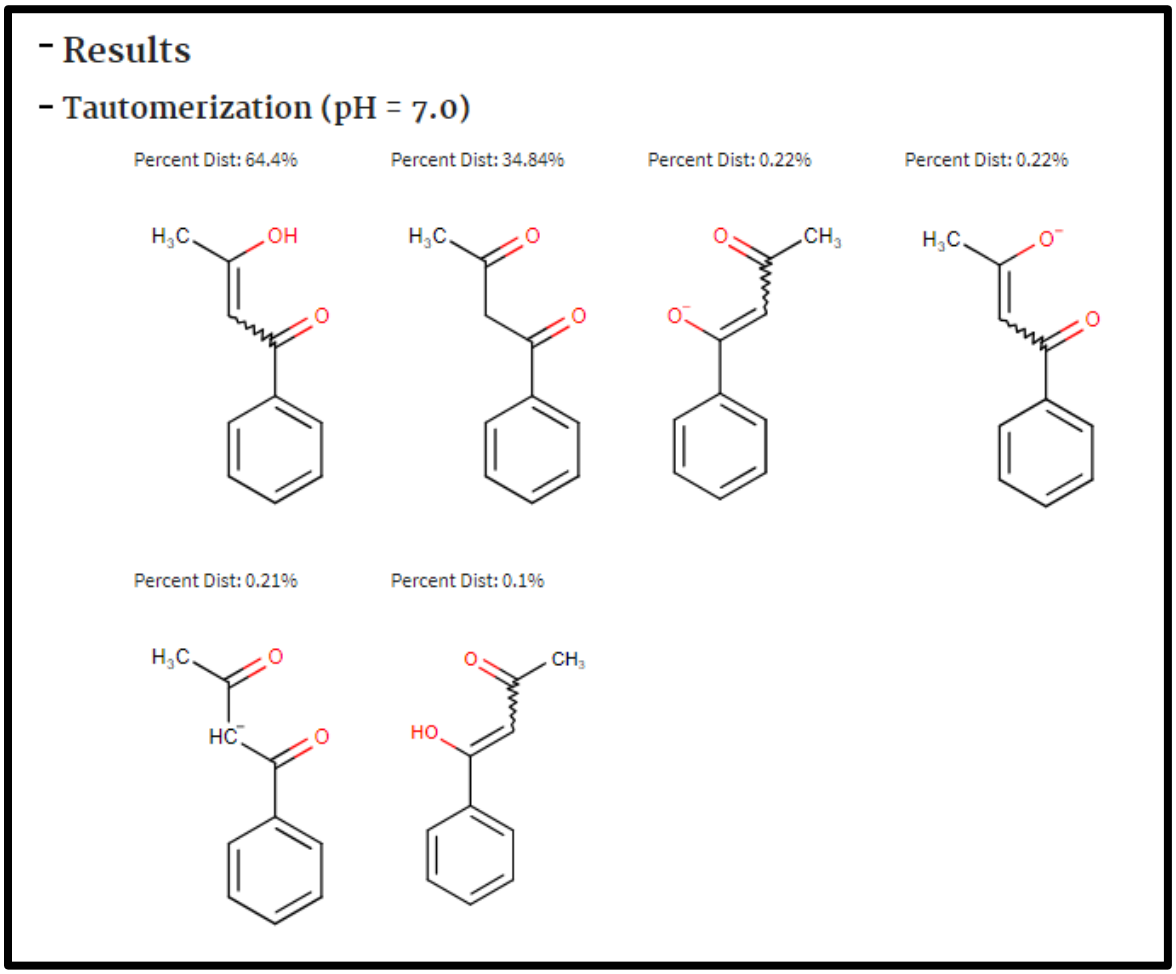


Clicking the Next button brings up the Calculate Chemical Speciation Workflow Inputs page. After selecting the Calculate Dominant Tautomer Distribution option, enter a limit for the number of possible tautomers and the pH value for which the distribution will be calculated. The default values are pH 7.0 and a limit of 100 tautomers as shown in the screen shot below.



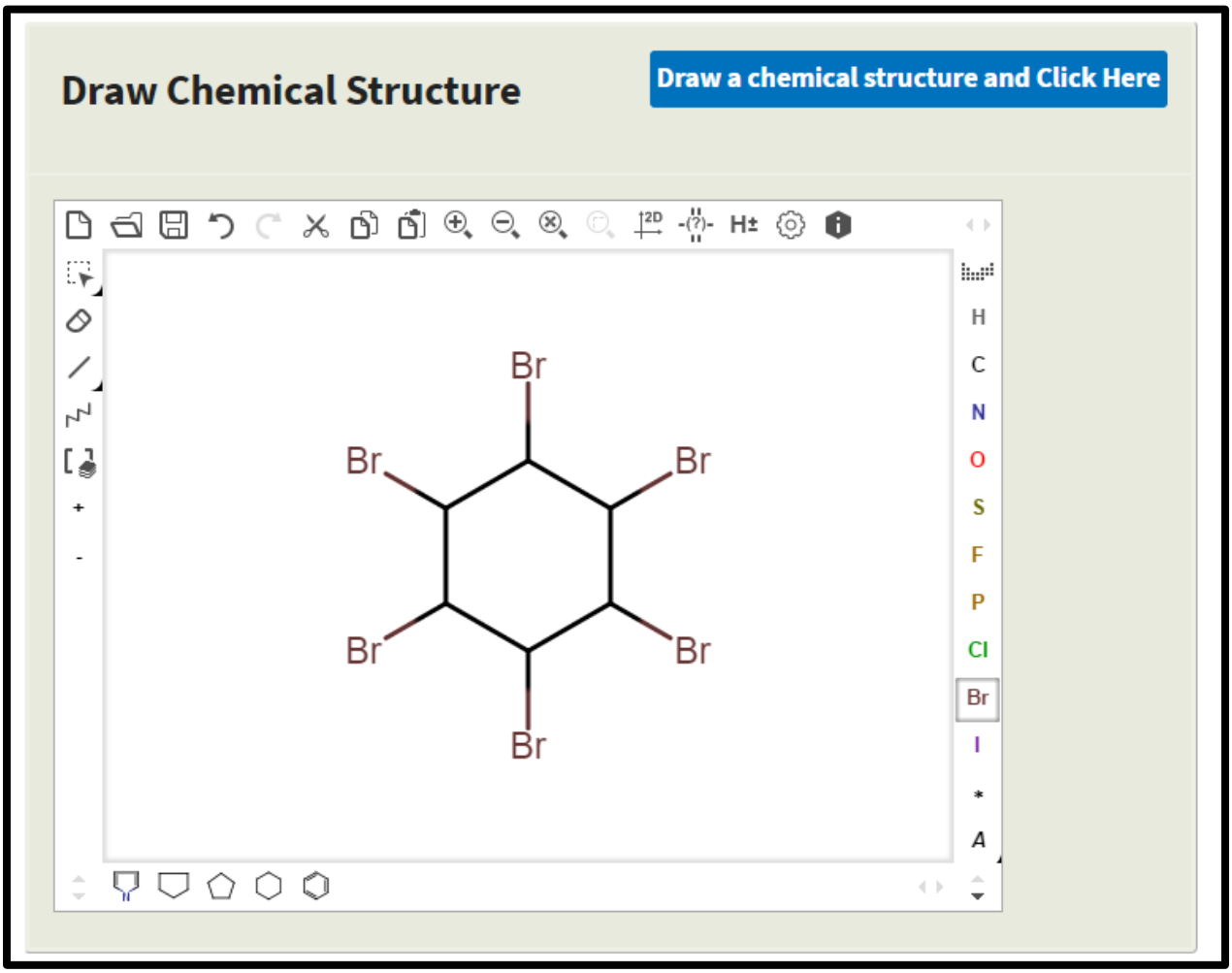
Click on the Submit button to view the output page for the tautomerization distribution based on the user-defined values.

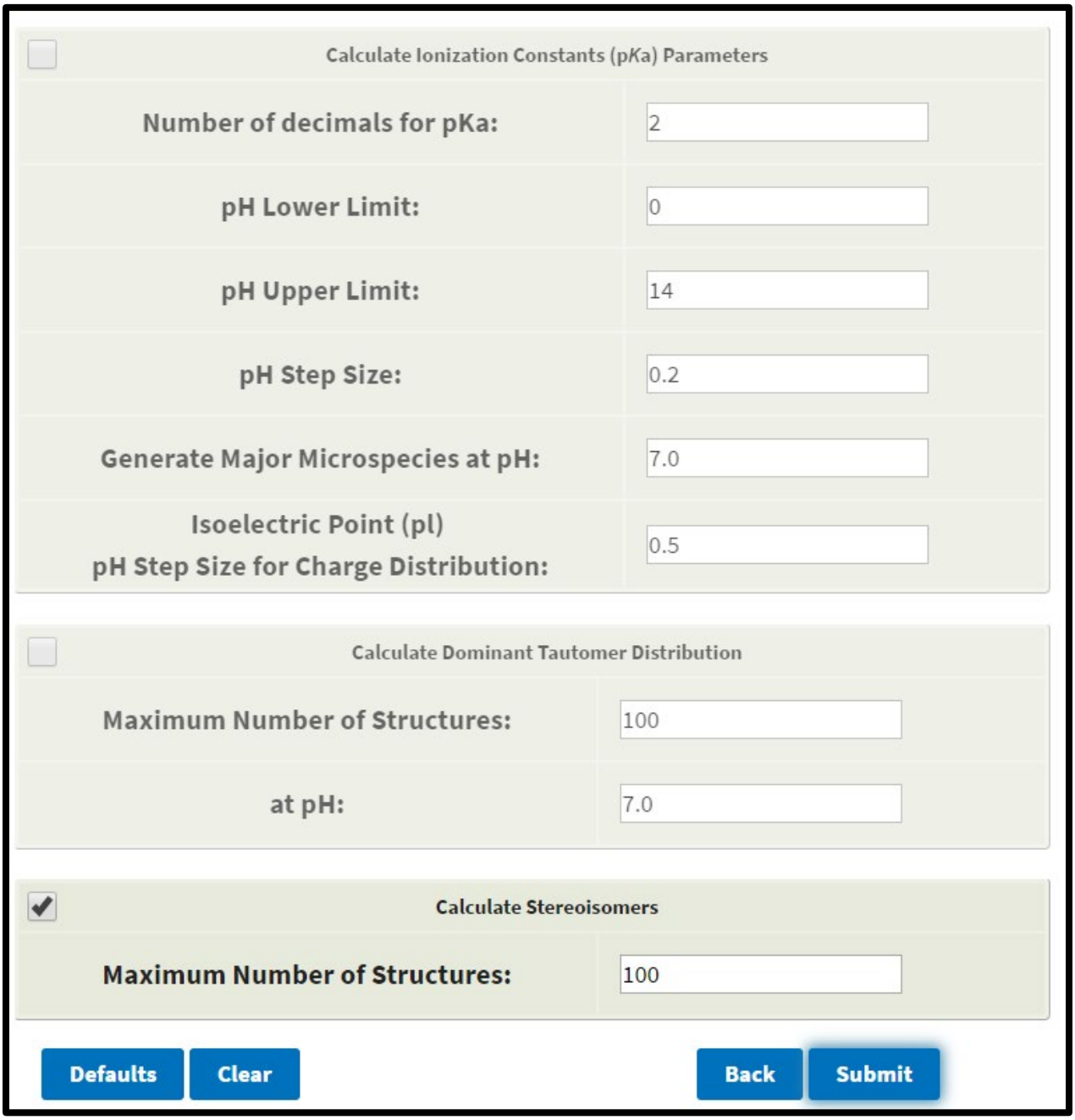
The Output screen shows the User Inputs (see screen shot above) as well as the tautomer distribution for the chemical of interest (see screen shot below). The individual structures can be enlarged by placing the cursor on top of the structure. The molecular information including the formula, IUPAC name, mass and SMILES string is also provided.



6.4. Calculate Stereoisomers

For this demonstration, 1,2,3,4,5,6-hexabromocyclohexane has been entered into the Chemical Editor as shown in the screen shot below.

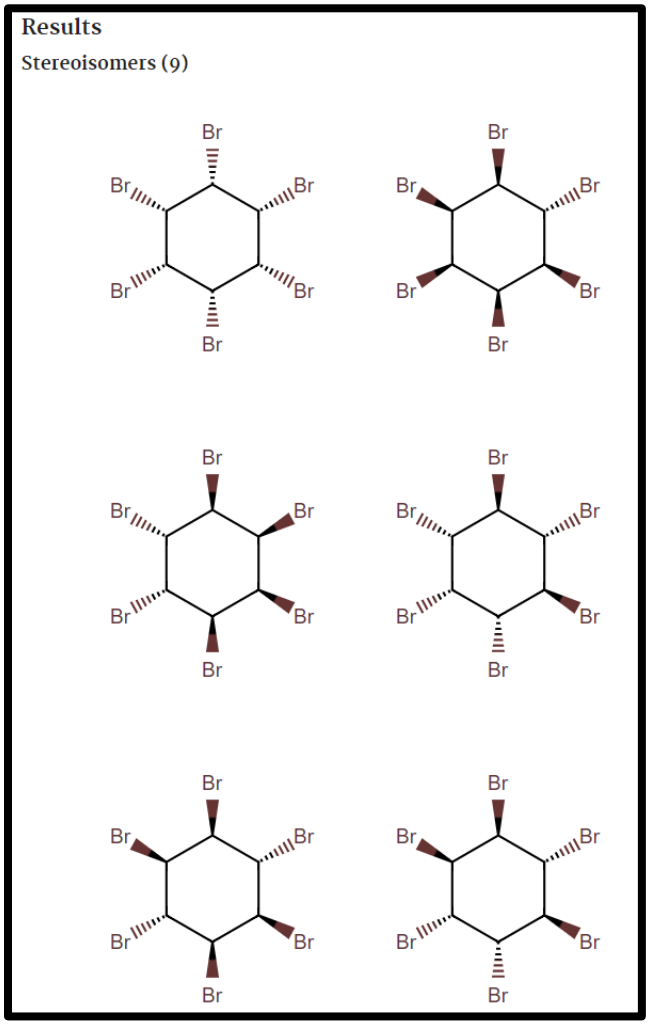


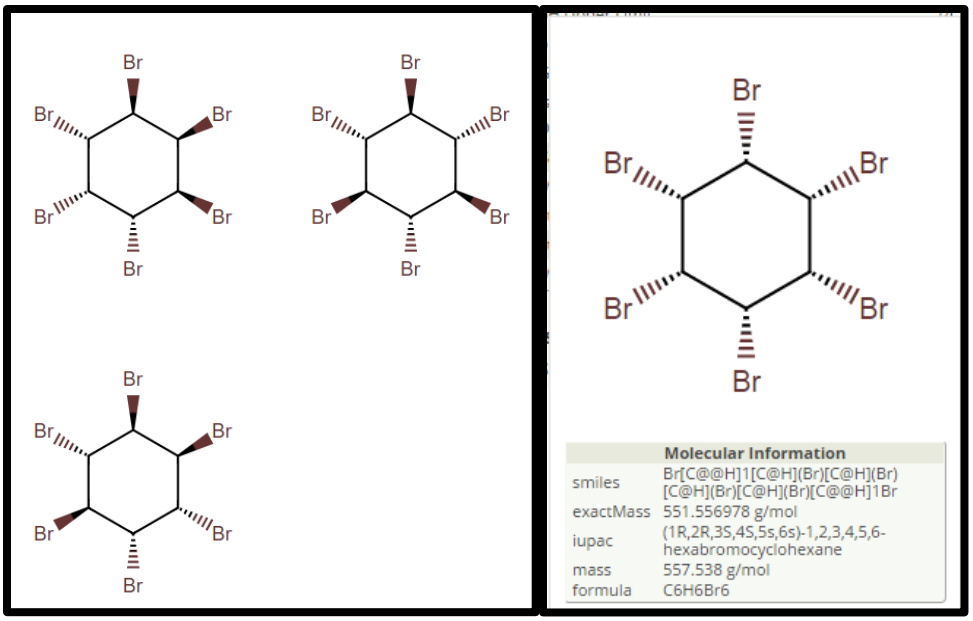


After selecting the Calculate Stereoisomers option, enter a limit for the maximum number of possible stereoisomers. The default value is 100 stereoisomers as shown in the screen shot above.

Clicking on the Next button provides the results of the calculation, which illustrate that 1,2,3,4,5,6-hexabromocyclohexane can exist as nine different stereo isomers as shown in the screen shot below. The individual structures can be enlarged by placing the cursor over the

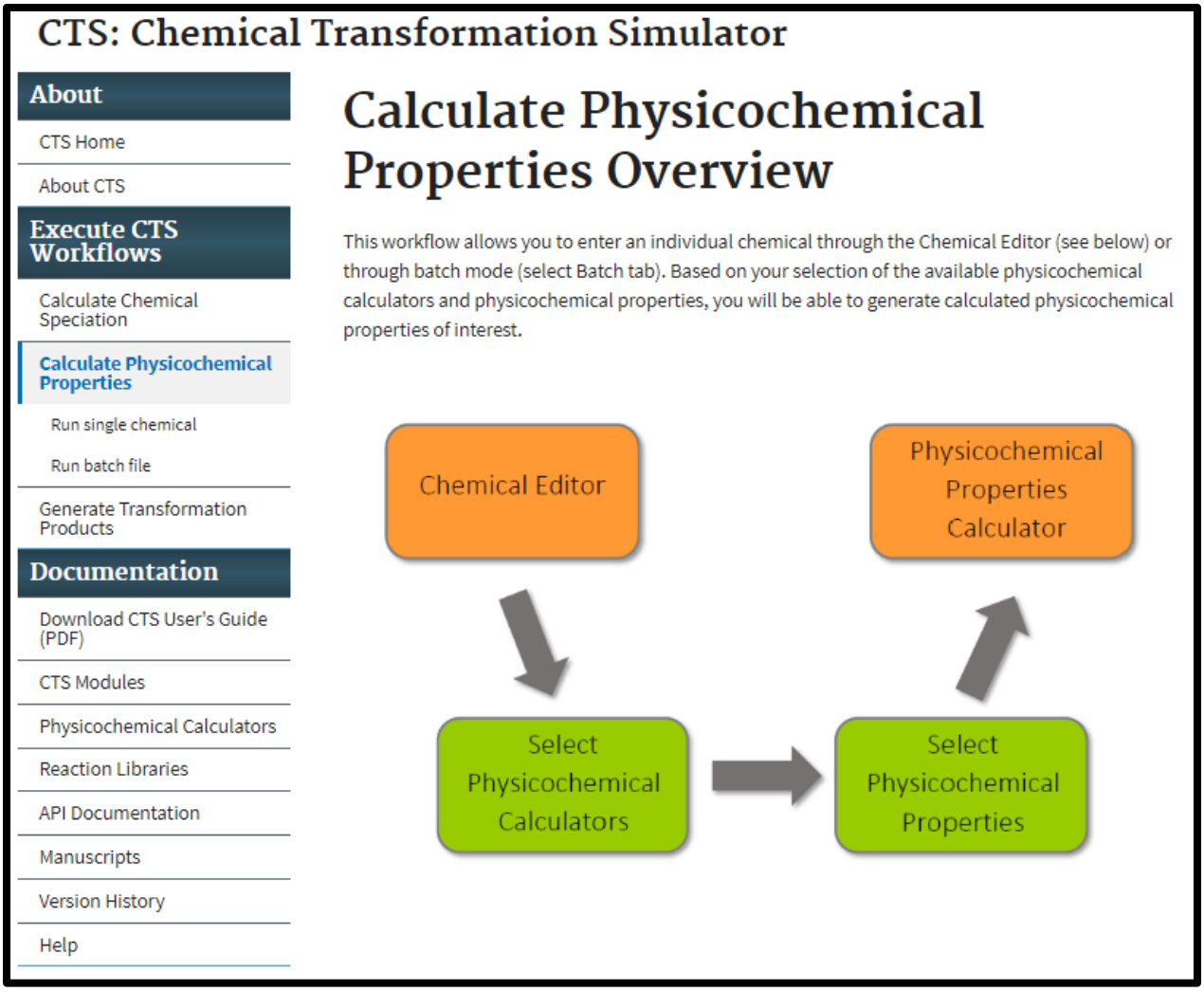
structure. The molecular information including the formula, IUPAC name, mass and SMILES string is also provided.



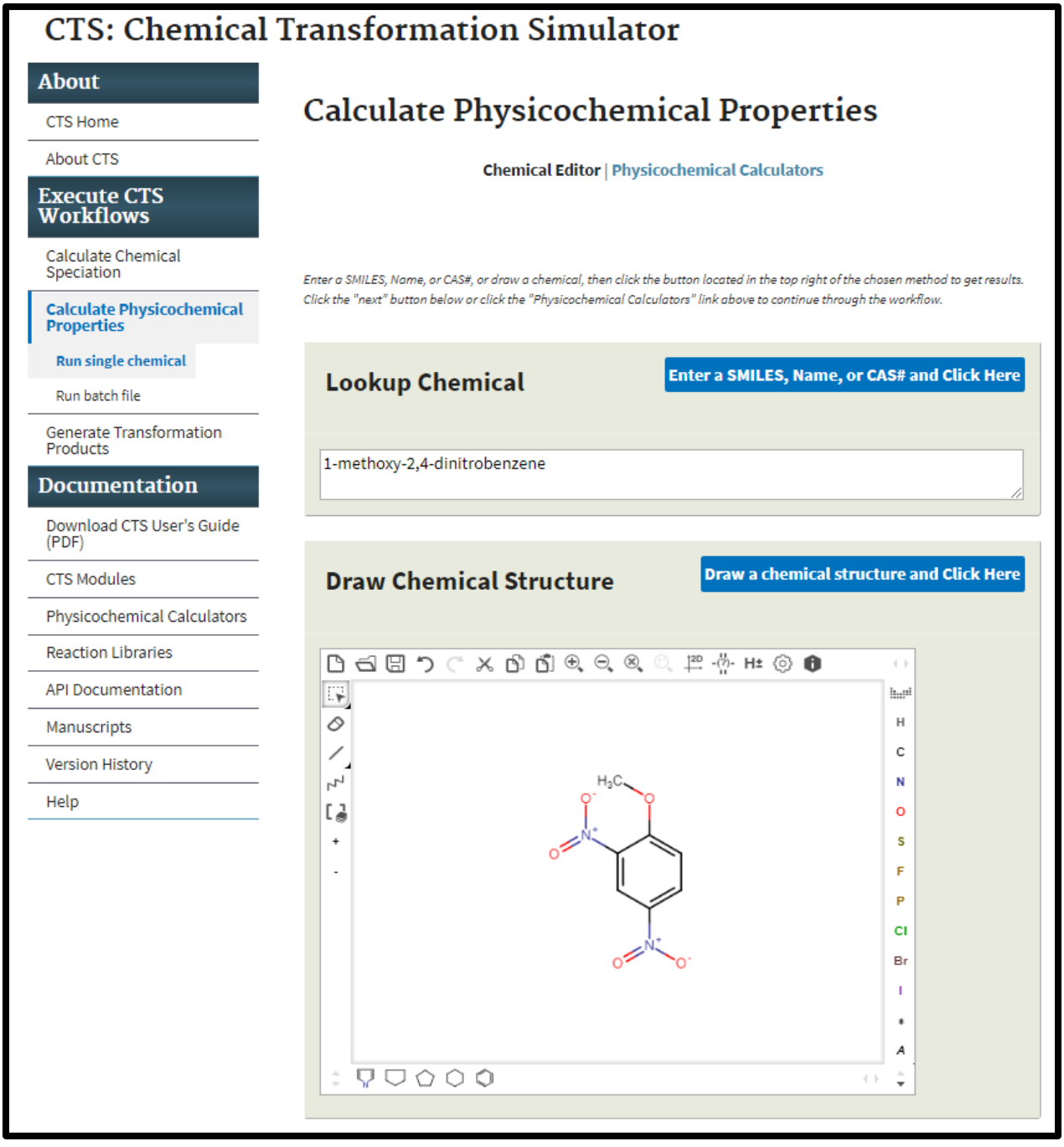

Br
Br
Br
Br
Br
Br
Br
Br
Br
Br
Br
Br
Br
Br
Br
Br
Br
Br
Molecular Information
smiles
Br[C@@H]1[C@H](Br)[C@H](Br)[C@H](Br)[C@H](Br)[C@@H]1Br
exactMass
551.556978 g/mol
iupac
(1R,2R,3S,4S,5s,6s)-1,2,3,4,5,6-hexabromocyclohexane
mass
557.538 g/mol
formula
C6H6Br6

6.5. Calculate Physicochemical Properties Workflow

Selection of the Calculate Physicochemical Properties Workflow provides the screen shot below illustrating the workflow overview. Click on the "Run single chemical" link to submit a single chemical for processing, or click on the "Run batch file" link to submit a batch file.



For this demonstration, 1-methoxy-2,4-dinitrobenzene has been entered into the Chemical Editor. The results are shown in the screen shots below. Select the Next button to choose the physiochemical calculators and physicochemical properties of interest.

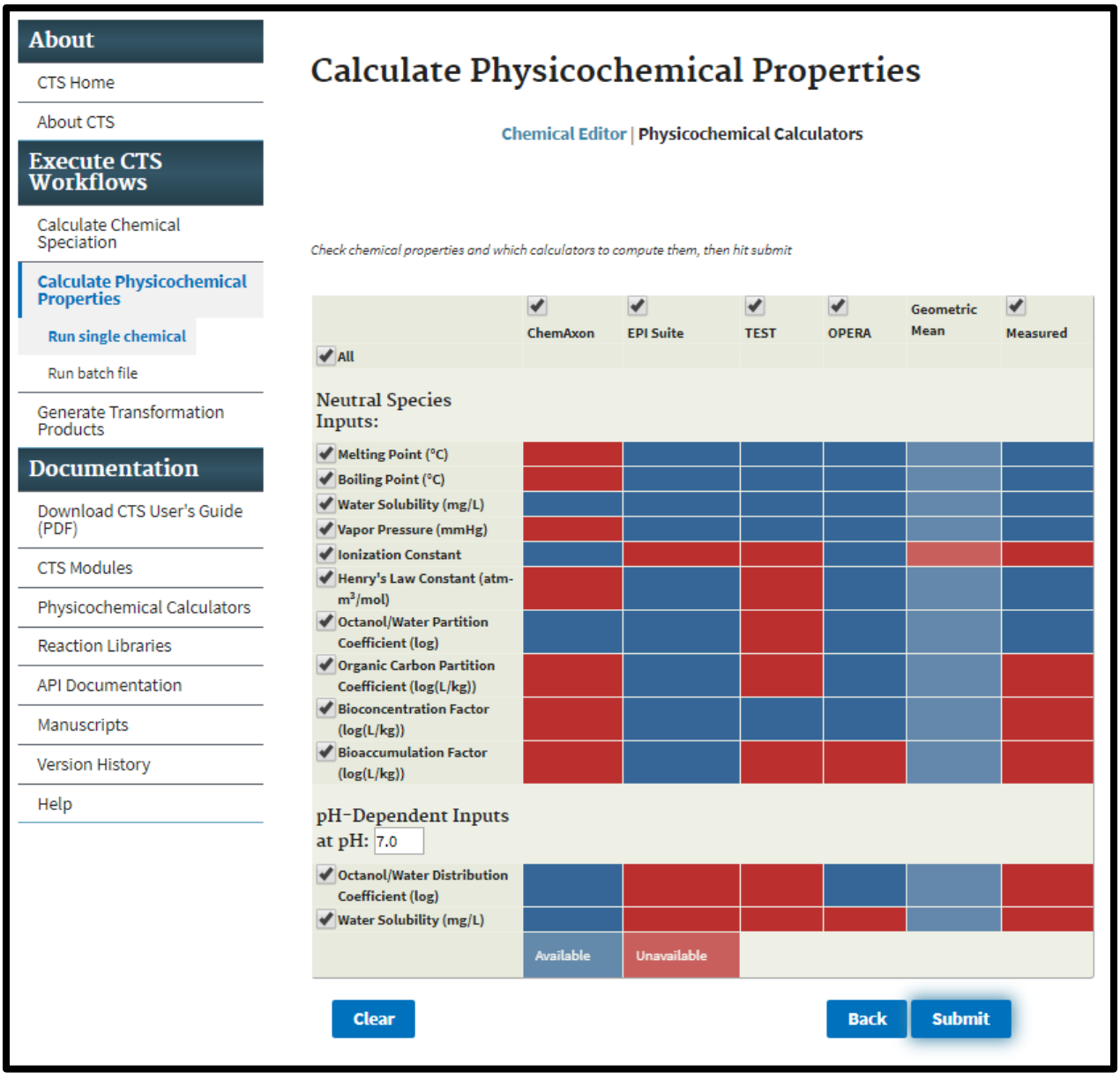


Results

Entered Chemical	1-methoxy-2,4-dinitrobenzene
Initial SMILES	COC1=CC=C(C=C1[N+]([O-])=O)[N+]([O-])=O
Standardized SMILES	COC1=CC=C(C=C1[N+]([O-])=O)[N+]([O-])=O
Preferred Name	2,4-Dinitroanisole
IUPAC	1-methoxy-2,4-dinitrobenzene
Formula	C7H6N2O5
Preferred CAS	119-27-7
Associated CAS	119-27-7
DTXSID	DTXSID9041366
Average Mass (g/mol)	198.134
Monoisotopic Mass (g/mol)	198.027671301

Clear Next

Use the Calculate Physicochemical Properties Workflow Inputs screen to select physicochemical properties and the physicochemical calculators of interest as shown in the screen shot below. Selection of the All button for the physicochemical properties selects all properties no matter which calculators are chosen, although only the available properties from each calculator will be calculated. See Table 1 for a summary of the calculators and calculation methods used in the Physicochemical Properties Calculator.



Application or Website	Version	Model	Property	Calculation Method	References
ChemAxon Plugin Calculators	16.10.17	KLOP	K_{ow}, D_{ow}	Group Contribution: MLR with fragment counts as descriptors	Klopman et al. (1994)
		VG	K_{ow}, D_{ow}	Group Contribution: MLR with fragment counts as descriptors	Viswanadhan et al. (1989)
		PHYS	K_{ow}, D_{ow}	Group Contribution: MLR with fragment counts as descriptors	Based on Viswanadhan et al. (1989) with PHYSPROP as training set
		Solubility Predictor	WS	Group Contribution: MLR with atom counts as descriptors	Hou et al. (2004)
		pKa Predictor	pK_a	Ionization site-specific regression equations	Szegezdi and Csizmadia (2004, 2007)
EPI Suite™	4.11	KOWWIN™	K_{ow}	Group Contribution: MLR with fragment counts as descriptors	Meylan and Howard (1995)
		WATERNT	WS	Group Contribution: MLR with fragment counts as descriptors	US EPA (2012)
		WSKOW	WS	MLR with log Kow, MP and MW as descriptors	Meylan et al. (1996); US EPA (2012)
		KOCWIN	K_{oc}	MCI-based QSAR; MLR with log Kow as descriptor	Meylan et al. (1992)
		MPBPVP	MP, BP, VP	Group Contribution: MLR with fragment counts as descriptors for MP and BP; VP from nonlinear function of BP	US EPA (2012)
		HENRYWIN	HLC	Bond Contribution: MLR with bond counts as descriptors	US EPA (2012)
		BCFBAF™	BCF and BAF	MLR with log Kow as descriptor; Arnot-Gobas method using upper trophic values	US EPA (2012)
T.E.S.T.	5.1.2†	Single Model	MP, BP, WS, VP, BCF	Hierarchical Clustering with similar chemicals	U.S. EPA (2020)
		Group Contribution	MP, BP, WS, VP, BCF	Group Contribution: MLR with fragment counts as descriptors	Martin and Young (2001)
		Hierarchical Clustering	MP, BP, WS, VP, BCF	Hierarchical Clustering	Martin et al. (2008)
		Nearest Neighbor	MP, BP, WS, VP, BCF	Average property value for 3 most similar molecules based on cosine similarity coefficient	Martin et al. (2008); U.S. EPA (2020)
OPERA	2.9‡	N/A	MP, BP, WS, VP, pK_a, HLC, K_{ow}, K_{oc}, BCF, D_{ow}	Weighted k-nearest neighbor using 2D descriptors from PaDEL	Mansouri et al. (2018)

Table 1. Summary of the calculators and calculation methods used in the Physicochemical Properties module.

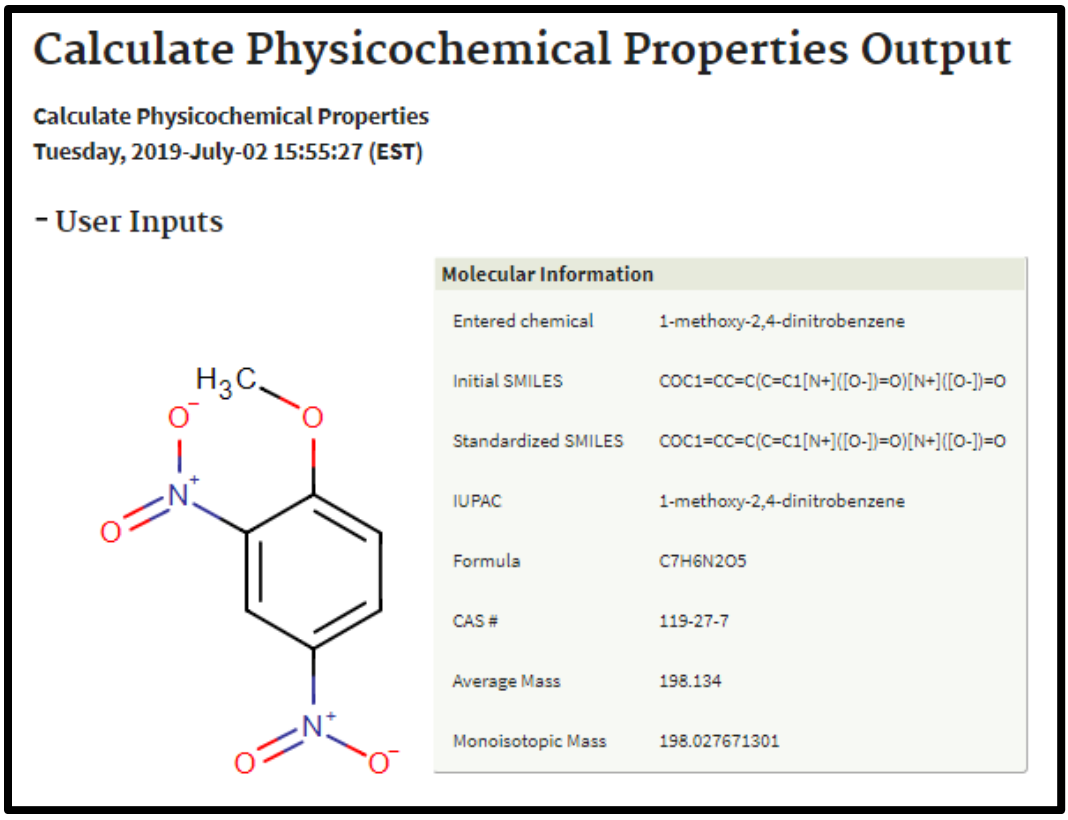
†T.E.S.T. calculations are retrieved through webservices from EPA's CompTox Chemistry Dashboard. The values shown on CTS reflect the current version of T.E.S.T. on the Dashboard.

‡OPERA values are retrieved through either a database of pre-calculated values or calculated on-the-fly, depending on the chemical. See below for details.

After selection of the physicochemical properties and calculators, selection of the Calculate data button provides the physicochemical properties output as illustrated in the screen shots below. Selection of the Measured checkbox provides available experimental data from the PHYSPROP database.

A database of pre-calculated values from the OPERA calculator for several thousand common chemicals has been implemented to improve the speed of retrieval from OPERA. If the given chemical is not in the pre-calculated database, the OPERA calculator will run the models on-the-fly, causing slower retrieval times for OPERA values. OPERA values for pH-dependent octanol-water partition coefficient (logD, also called D_{ow} or d_ow) are only available at pH 5.5 and 7.4.

The geometric mean of the predicted values (not including any measured data or ionization constants) for each selected property from the selected calculators will be automatically calculated and displayed under the Geometric Mean column when predicted values are requested. For some chemicals, the value of the melting point and/or boiling point predicted by one or more of the calculators may be a negative value in units of Celsius. Therefore, to calculate the geometric mean of melting point and boiling point, the calculated values are converted from degrees Celsius to Kelvin before calculating the geometric mean and converting that value back to degrees Celsius.



Calculate Physicochemical Properties Output

Calculate Physicochemical Properties
Tuesday, 2019-July-02 15:55:27 (EST)

- User Inputs

Molecular Information	
Entered chemical	1-methoxy-2,4-dinitrobenzene
Initial SMILES	COC1=CC=C(C=C1[N+]([O-])=O)[N+]([O-])=O
Standardized SMILES	COC1=CC=C(C=C1[N+]([O-])=O)[N+]([O-])=O
IUPAC	1-methoxy-2,4-dinitrobenzene
Formula	C7H6N2O5
CAS #	119-27-7
Average Mass	198.134
Monoisotopic Mass	198.027671301

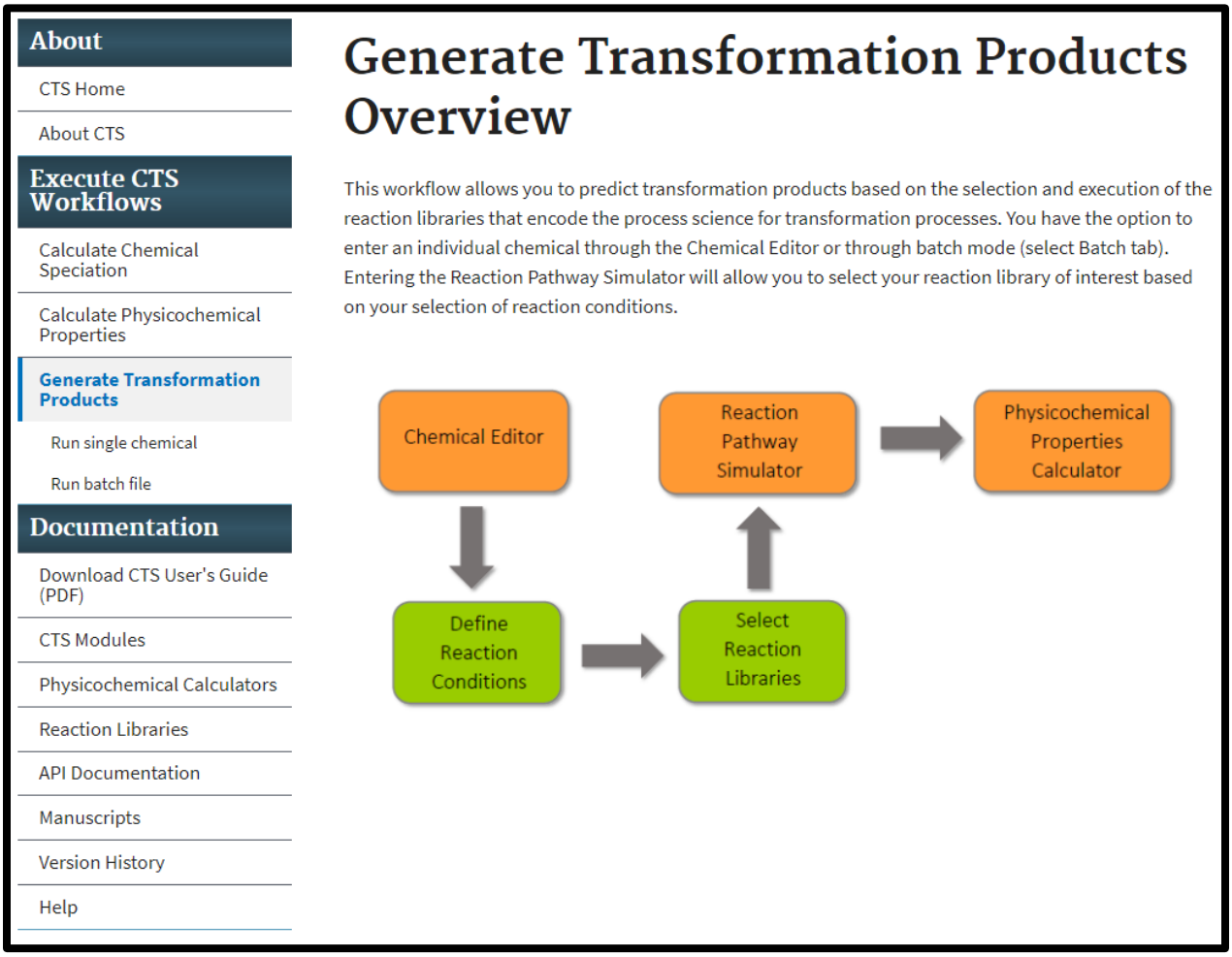
- Physicochemical Properties Results

All	ChemAxon	EPI Suite	TEST	OPERA	Geometric Mean	Measured
Neutral Species Inputs:						
Melting Point (°C)		96.56	78.20 HC 87.50 NN 80.30 GC	95.09	87.45	94.50
Boiling Point (°C)		319.62	286.50 HC 308.70 NN 315.00 GC	304.17	306.68	206.00
Water Solubility (mg/L)	2.68e+2	3.66e+2 WSKOW 2.58e+2 WATERNT	9.54e+1 HC 7.97e+1 NN 5.94e+2 GC	3.21e+2	2.33e+2	1.55e+2
Vapor Pressure (mmHg)		1.38e-4	2.06e-4 HC 1.14e-4 NN 2.17e-4 GC	3.05e-4	1.85e-4	N/A
Ionization Constant	none			pKa_1: 7.25		
Henry's Law Constant (atm-m^3/mol)		4.96e-9		5.52e-7	5.23e-8	N/A
Octanol/Water Partition Coefficient (log)	1.74 KLOP 1.70 VG 1.65 PHYS	1.71		1.94	1.75	N/A
Organic Carbon Partition Coefficient (log(L/kg))		2.36 MCI 2.37 KOW		2.53	2.42	
Bioconcentration Factor (log(L/kg))		0.80 REG 0.68 A-G	0.73 SM 0.10 HC 0.48 NN 0.96 GC	0.98	0.68	
Bioaccumulation Factor (log(L/kg))		0.68 A-G			0.68	
pH-Dependent Inputs at pH: 7.0						
Octanol/Water Distribution Coefficient (log)	1.74 KLOP 1.70 VG 1.65 PHYS			1.56	1.66	
Water Solubility (mg/L)	-2.87e+3				-2.87e+3	
	Available	Unavailable				

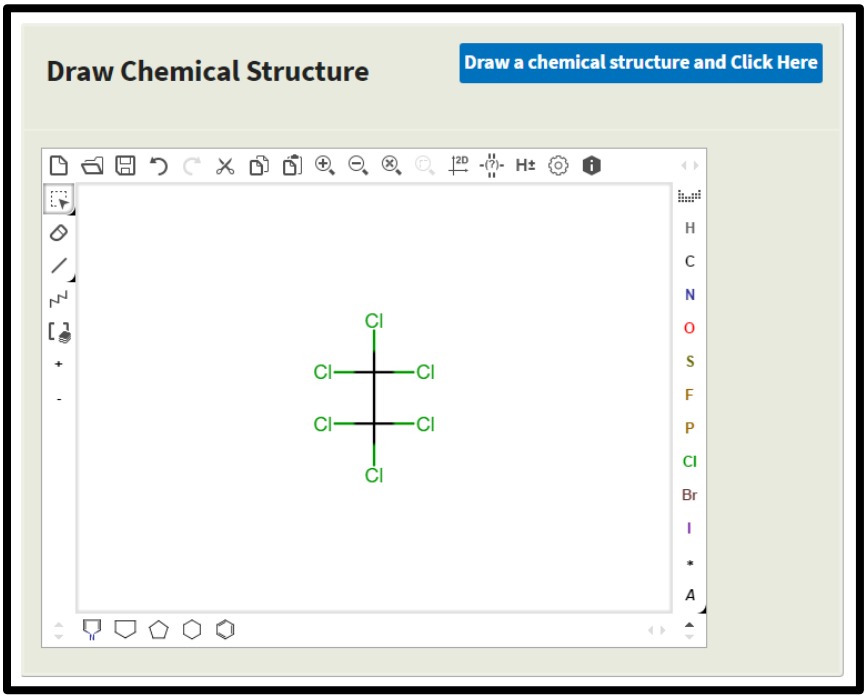
Cancel | Clear data | Calculate data

6.6. Generate Transformation Products Workflow

Selection of the Generate Transformation Products Workflow provides the screen shot below illustrating the workflow overview. Click on the "Run single chemical" link to submit a single chemical for processing, or click on the "Run batch file" link to submit a batch file.



For this demonstration, hexachloroethane has been entered into the Chemical Editor as illustrated in the screen shot below.



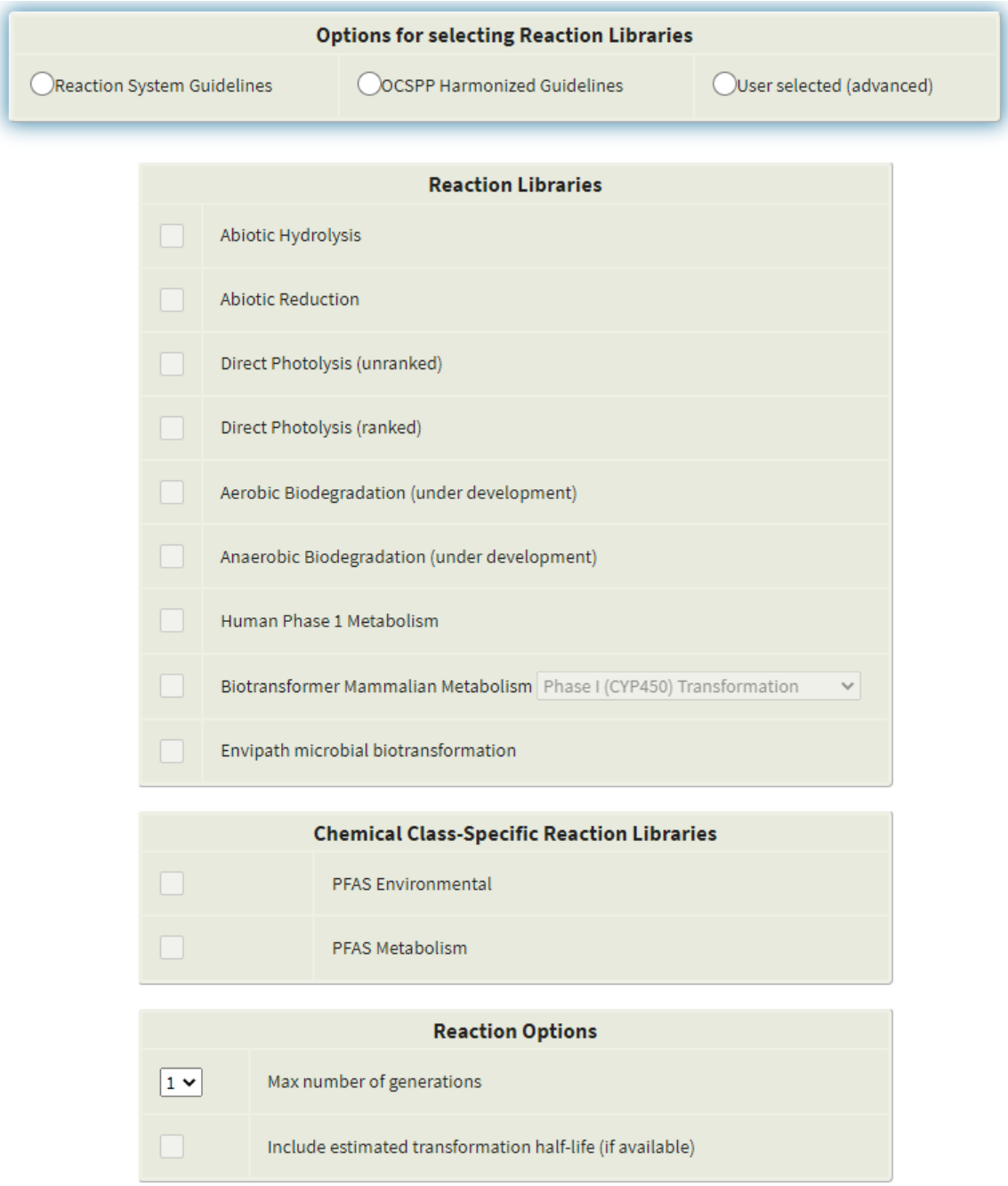
The first required input is the selection of one or more reaction libraries based on the transformation pathways of interest. Reaction libraries have been developed for the environmental transformation processes of abiotic hydrolysis, abiotic reduction, and direct photolysis (unranked and ranked). A Spontaneous reaction library, which is intended to be combined with other CTS reaction libraries, is also available to capture the rapid transformation of intermediates that form due to other transformation processes (e.g., hydrolysis or photolysis). A reaction library for human phase I metabolism that was developed by ChemAxon is also available through the RPS. Additionally, linkages are available to the EnviPath prediction system for environmental microbial biodegradation pathways and to the BioTransformer tool, which includes four independent transformation modules for metabolite prediction: Phase I (CYP450) metabolism in mammals, Phase II (CYP450) conjugation reactions in mammals, gut microbial metabolism, and Enzyme Commission (EC)-Based reactions. Finally, two chemical class-specific libraries have been developed to predict the likely environmental and metabolic transformation products of per- and polyfluoroalkyl substances (PFAS).

There are three available options for the selection of one or multiple reaction libraries (see screen shot below):

- Reaction System Conditions
- OCSPP Harmonized Guidelines
- User Selected (Advanced)

The OCSPP Harmonized Guidelines specify EPA/OECD-recommended methods to generate data that is submitted to EPA to support:

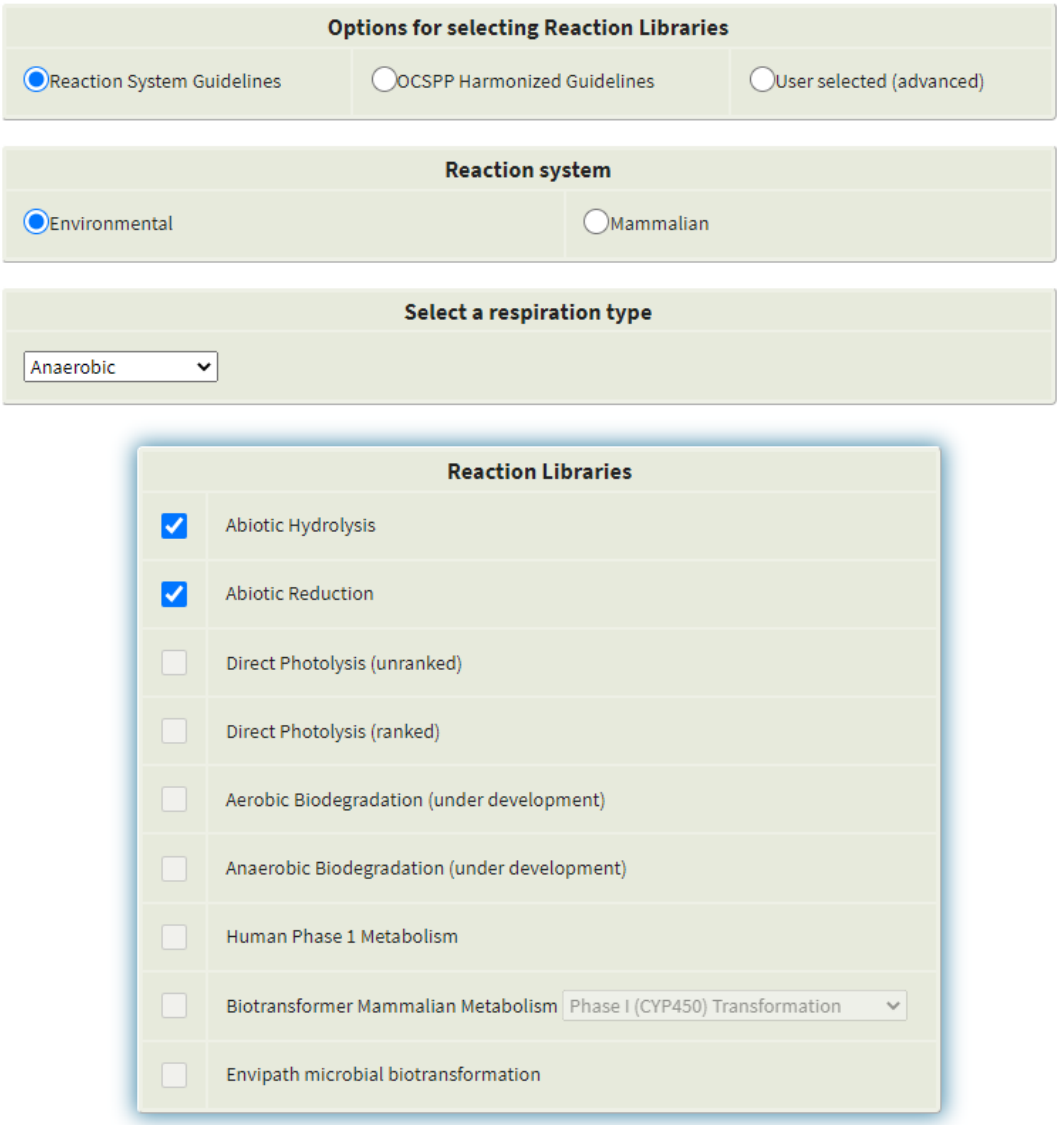
- The registration of a pesticide under the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA);
- The decision-making process supporting potential regulation of an industrial chemical under the Toxic Substances Control Act (TSCA)



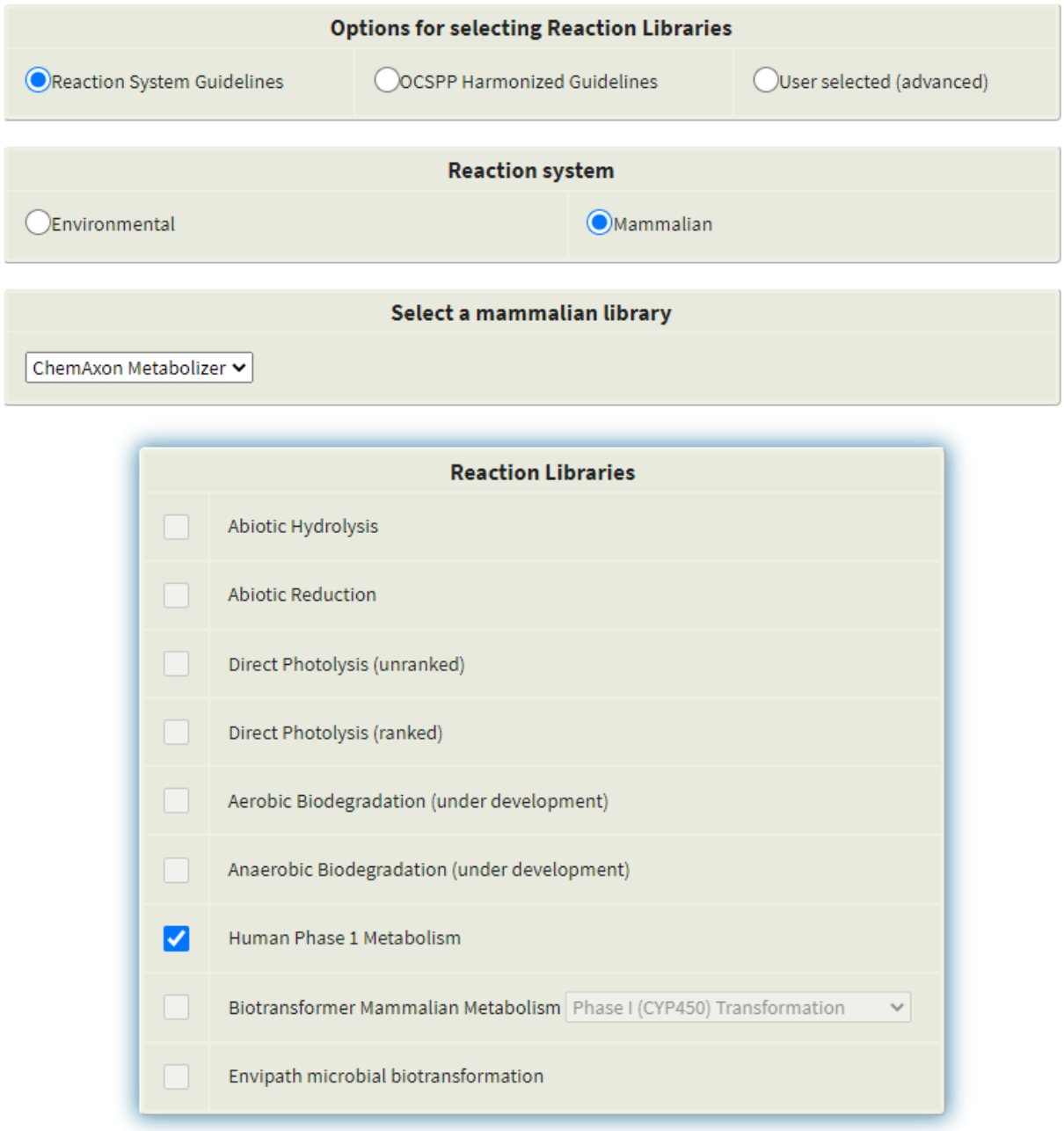
6.6.1 Reaction Library Selection

Selection of reaction libraries through the Reaction System Conditions option provides two reaction system types: Environmental or Mammalian.

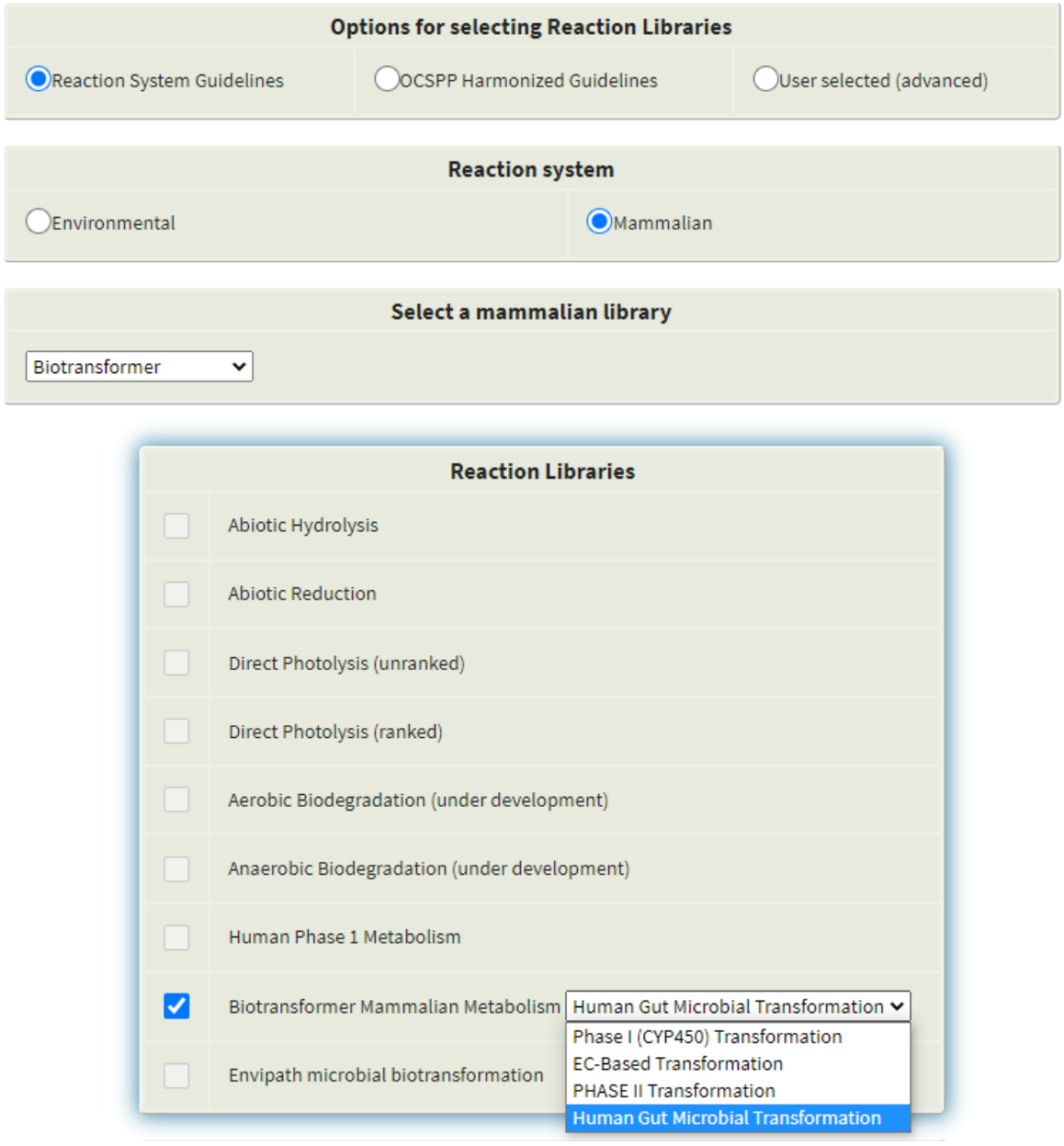
Under the Environmental Reaction System option, three respiration types are available: Aerobic (abiotic), Aerobic (biotic), or Anaerobic. When the respiration type is selected, the system identifies the relevant reaction libraries. For the Aerobic (abiotic) respiration type, the relevant libraries are abiotic hydrolysis and direct photolysis. For the Aerobic (biotic) reaction type, predicted products are obtained from the EnviPath prediction system for environmental microbial biodegradation pathways. For the Anaerobic respiration type, the relevant reaction libraries are abiotic hydrolysis and abiotic reduction (see screen shot below).



Under the Mammalian Reaction System option, user has the choice of selecting between ChemAxon Metabolizer or Biotransformer libraries. Selection of ChemAxon Metabolizer opens the window with the human phase I metabolism reaction library selected as shown in the screenshot below.



Selection of the Biotransformer option opens the window with Biotransformer mammalian metabolism selected, prompting the user to select one from four library options: Human Gut Microbial Transformation, PHASE II Transformation, EC-Based Transformation, and Phase I (CYP450) Transformation.

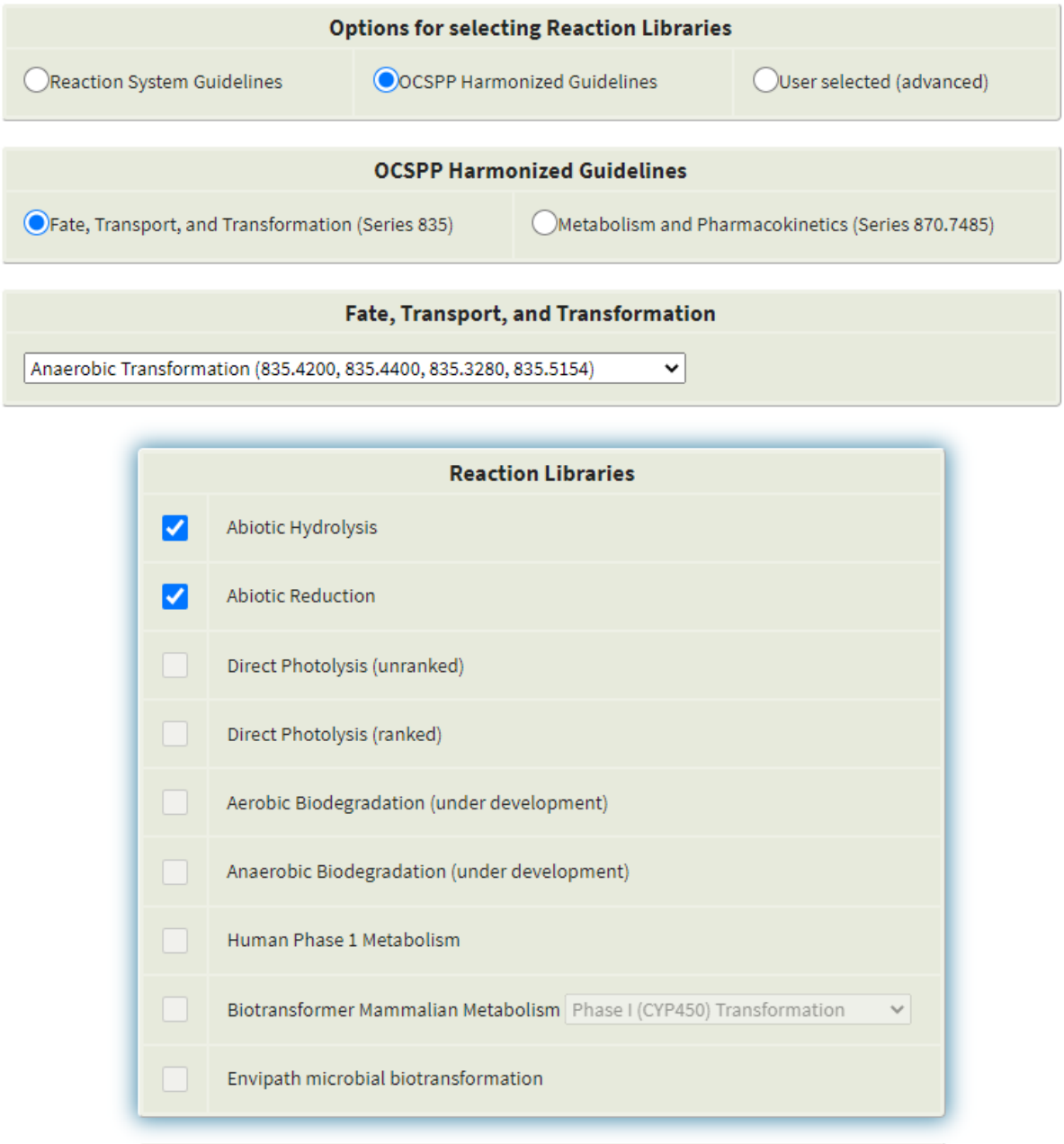


The second option for the selection of reaction libraries is through the selection of the Office of Chemical Safety and Pollution Prevention (OCSPP) Fate, Transport, and Transformation (Series 835) or Metabolism and Pharmacokinetics (Series 870.7485) Guidelines.

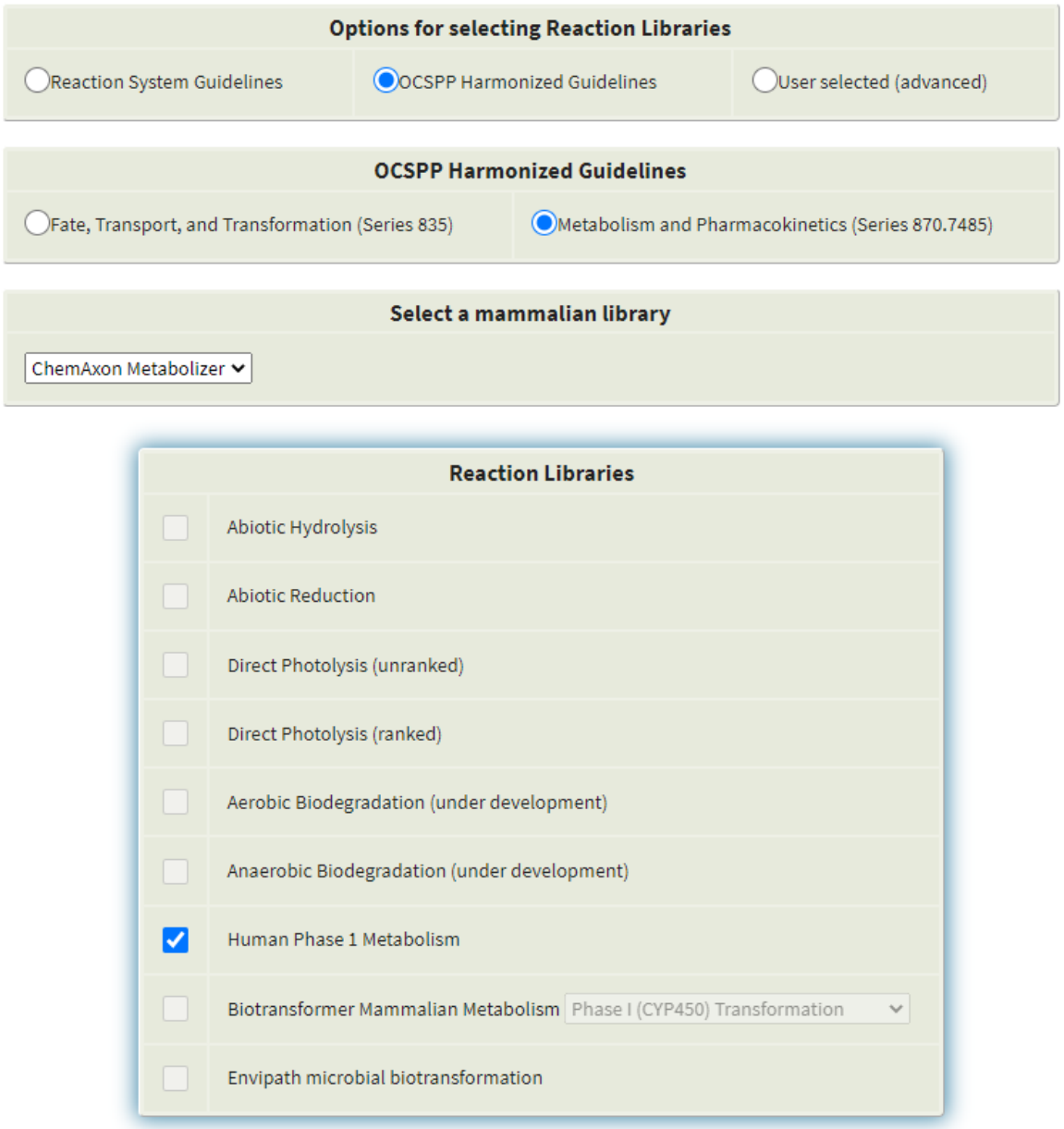
Selection of the Fate, Transformation Series provides three options:

- Hydrolysis (835.2120 or 835.2240)
- Direct Photolysis (835.2210 or 835.2240)
- Aerobic Transformation (835.4100, 835.3180, 835.3190, 835.3280)
- Anaerobic Transformation (835.4200, 835.4400, 835.3280, 835.5154)

As an example, selection of the Anaerobic Transformation (835.4200, 835.4400, 835.3280, 835.5154) shows that both the abiotic hydrolysis and abiotic reduction are appropriate selections for this option as shown in the screen shot below.



Selection of Metabolism and Pharmacokinetics (Series 870.7485) provides two options: ChemAxon Metabolizer or Biotransformer. Selection of ChemAxon Metabolizer provides one option for selection of a reaction library (i.e., Human Phase I Metabolism) as shown in the screen shot below.



The third option for the selection of reaction libraries is through the selection of the "User selected" button as shown in the screen shot below. This option provides the ability to select amongst the currently available reaction libraries. Along with the libraries based on

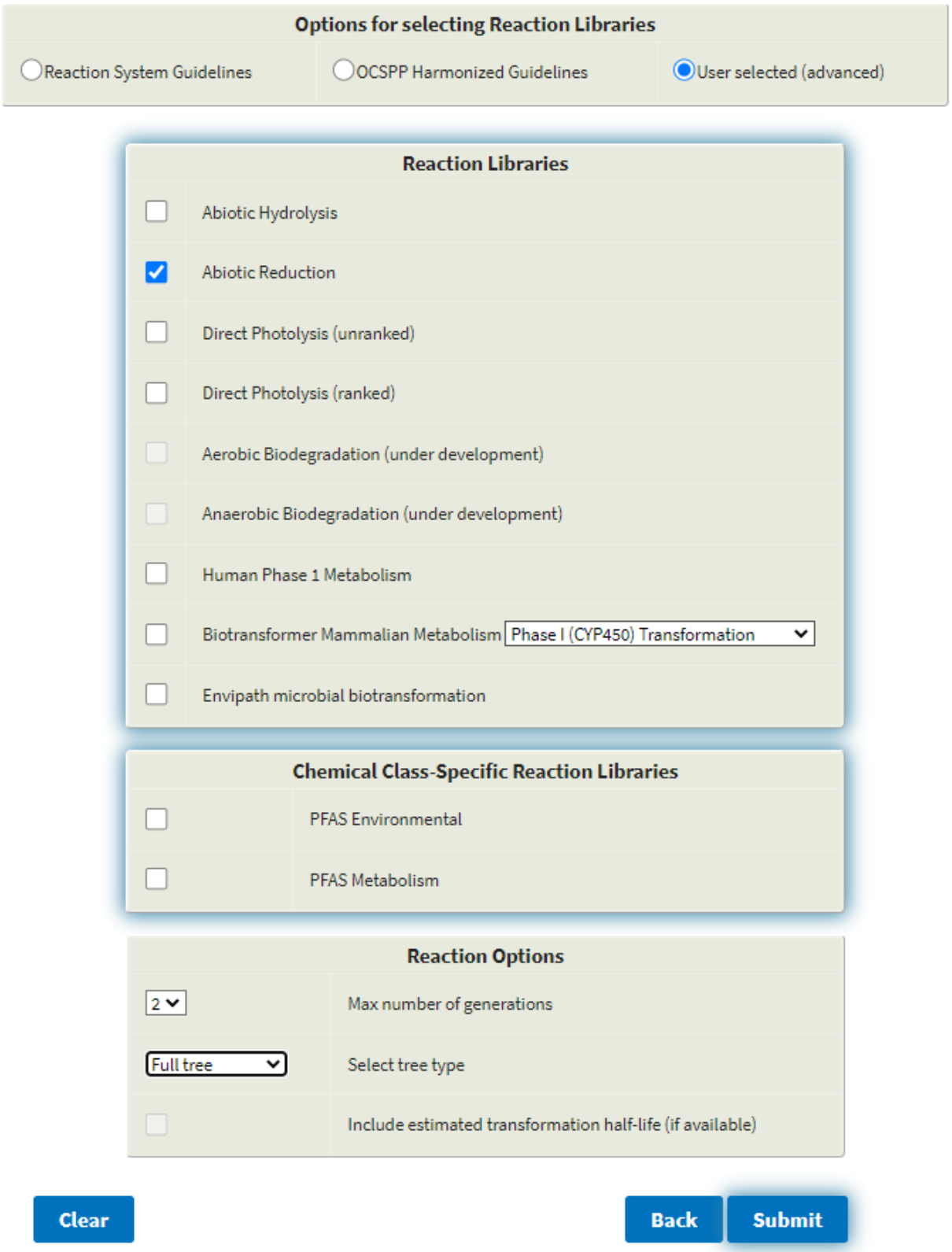
transformation type, class-specific libraries are available for PFAS compounds undergoing environmental transformation or metabolism.

After reaction library selection, three additional options are available in the Reaction Options box:

- **Max number of generations:** A dropdown list allows the user to select the maximum number of generations of transformation products that will be generated. The default value is set at one, and the maximum value is four for most libraries. For the Direct Photolysis libraries, the maximum value is two to avoid producing an excess of products.
- **Select tree type**: a dropdown list allows the user to select either a "Simplified Tree" or a "Full Tree" option. When the "Simplified Tree" option is selected, the resulting transformation product tree will show only one instance of a specific parent/product relationship within a single generation. This means that when a particular product can be formed from a parent by more than reaction scheme in a single generation, the "Simplified Tree" will display that product only once, rather than showing the formation of the product from the same parent multiple times. However, when a product can be formed within a subsequent generation or from a different parent within the same generation, multiple copies of the product will be shown in the tree. This output is helpful for those who want a cleaner view of transformation products while looking for a specific product, or for users interested in collecting unique products. When "Full Tree" option is selected, the resulting transformation product tree will display all transformation pathways with all products.
- **Include estimated transformation half-life (if available)**: When this box is checked, CTS will provide estimated reaction scheme-specific transformation half-life values for the chemical of interest and transformation products. Currently, the only available estimated half-life values are for selected schemes in the abiotic hydrolysis reaction library. Schemes outside of this scope will report the default qualitative descriptor instead (see section 6.6.2, page 49). The estimated half-life values are obtained from the EPI Suite™ HYDROWIN module, which calculates hydrolysis rate constants using fragment-based QSAR models (Drossman et al., 1988). The table below provides a crosswalk between the relevant schemes within the CTS Hydrolysis Reaction library and the corresponding HYDROWIN model. Importantly, the option to show estimated transformation half-life values is only available when viewing the results as Full Tree output.

CTS Hydrolysis Scheme Name	HYDROWIN Model
Halogenated Aliphatics: Nucleophilic Substitution (no adjacent X)	HALOMETHANE or ALKYL HALIDE
Halogenated Aliphatics: Nucleophilic Substitution (vicinal X)	ALKYL HALIDE

Halogenated Aliphatics: Nucleophilic Substitution (geminal X)	ALKYL HALIDE
Halogenated Aliphatics: Elimination	ALKYL HALIDE
Epoxide Hydrolysis	EPOXIDE
Organophosphorus Ester Hydrolysis 1	Thiophosphate/Phosphate
Organophosphorus Ester Hydrolysis 2	Thiophosphate/Phosphate
Carboxylic Acid Ester Hydrolysis	ESTER
Anhydride Hydrolysis	ESTER
Carbamate Hydrolysis	CARBAMATE


Options for selecting Reaction Libraries
Reaction System Guidelines
OCSPP Harmonized Guidelines
User selected (advanced)
Reaction Libraries
Abiotic Hydrolysis
Abiotic Reduction
Direct Photolysis (unranked)
Direct Photolysis (ranked)
Aerobic Biodegradation (under development)
Anaerobic Biodegradation (under development)
Human Phase 1 Metabolism
Biotransformer Mammalian Metabolism
Phase I (CYP450) Transformation
Envipath microbial biotransformation
Chemical Class-Specific Reaction Libraries
PFAS Environmental
PFAS Metabolism
Reaction Options
2
Max number of generations
Full tree
Select tree type
Include estimated transformation half-life (if available)
Clear
Back
Submit

6.6.2 Output from the Generate Transformation Products Workflow

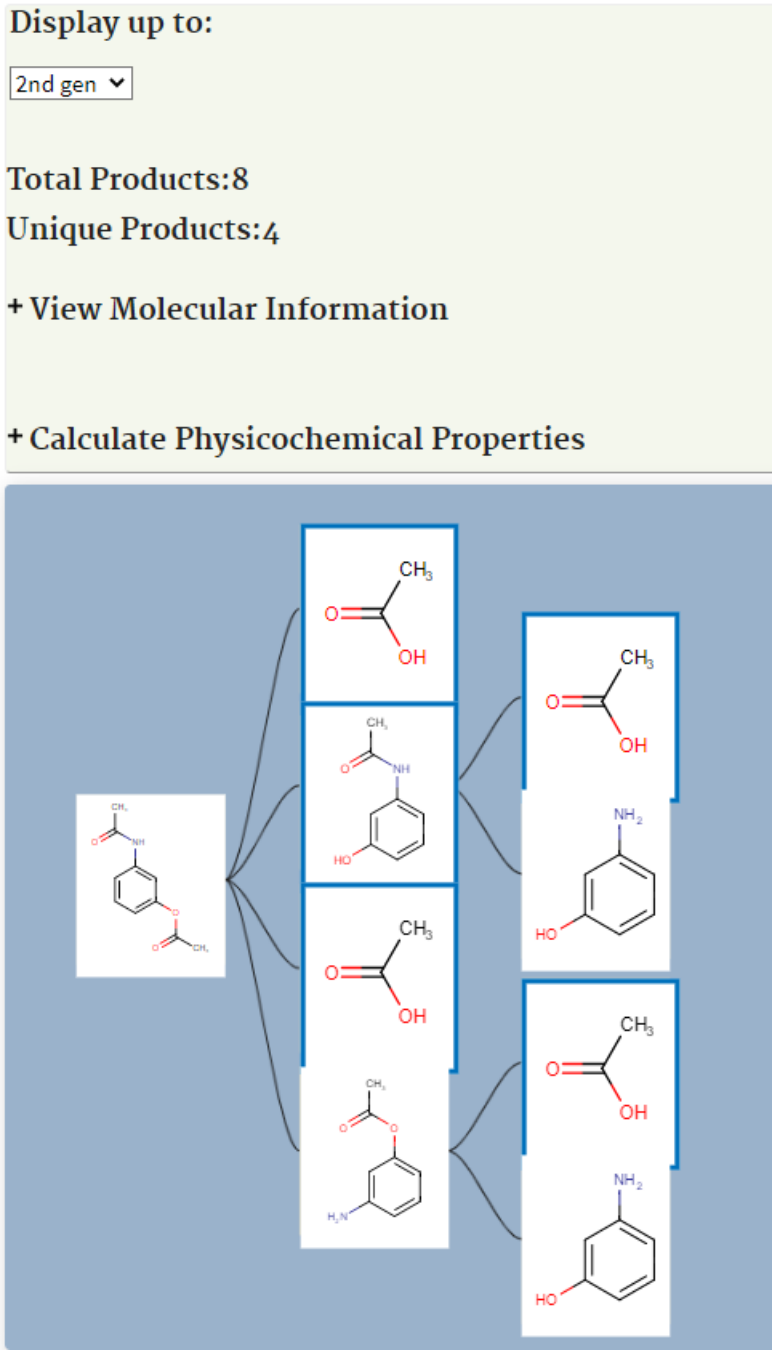
After selection of the reaction libraries and reaction options have been made, click the Submit button to generate transformation products. The screen shot below shows the upper portion of the output for the Generate Transformation Products workflow, which summarizes the input data and indicates the total number of products that are predicted. If "Total Products" is equal to zero, then the user is to assume the selected reaction processes do not occur for the parent chemical of interest.

For this demonstration, 3-acetamidophenyl acetate is the chemical of interest, and the selected reaction library is abiotic hydrolysis.

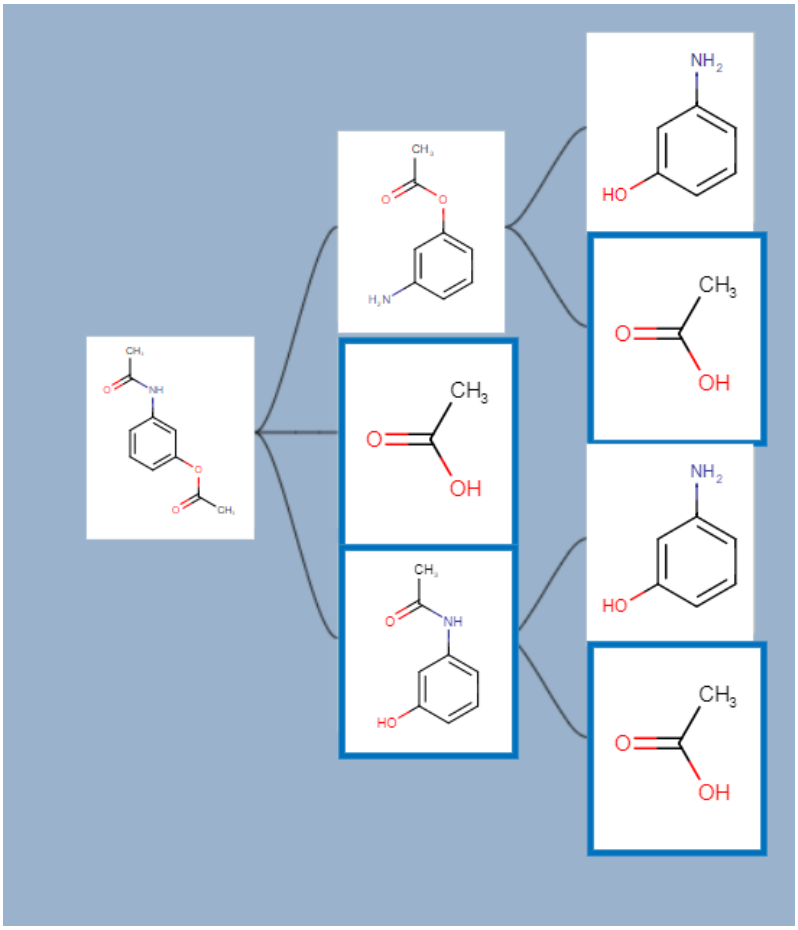
- User Inputs

Molecular Information	
Entered chemical	3-acetamidophenyl acetate
Initial SMILES	CC(=O)NC1=CC(OC(C)=O)=CC=C1
Standardized SMILES	CC(=O)NC1=CC(OC(C)=O)=CC=C1
IUPAC	3-acetamidophenyl acetate
Formula	C10H11NO3
CAS #	6317-89-1
Average Mass	193.202 g/mol
Monoisotopic Mass	193.073893218 g/mol
Reaction Pathway Simulator	
Libraries	hydrolysis
Generation Limit	2
Include rates	none
Tree type	full_tree

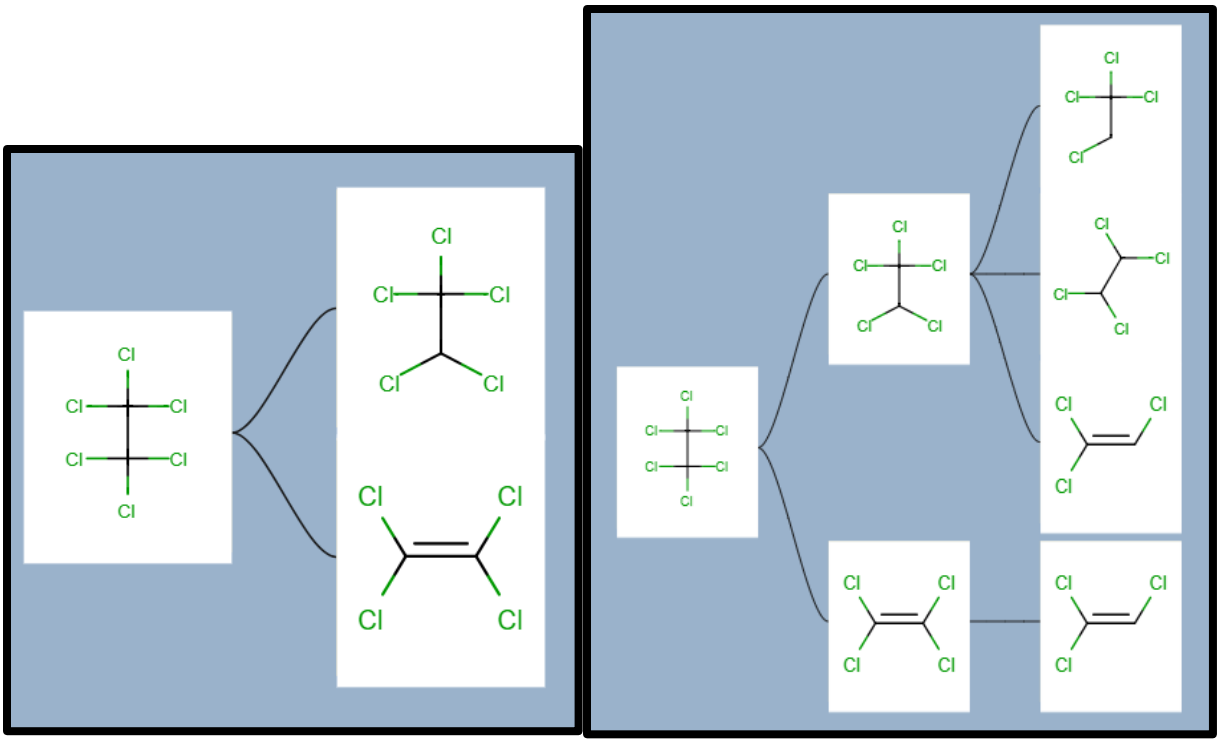
The lower portion of the output for the Generate Transformation Products workflow shows a reaction pathway map, and a dropdown menu can be used to select the number of generations displayed. The reaction pathway map displayed will either be the “Simplified Tree” or “Full Tree” that was selected by the user. In this demonstration, the “Full Tree” option was selected, so all predicted products are displayed.



If the Simplified Tree option was selected, replicate products are removed when they are formed within the same generation and from the same parent molecule; therefore, the number of displayed products per generation can be reduced.

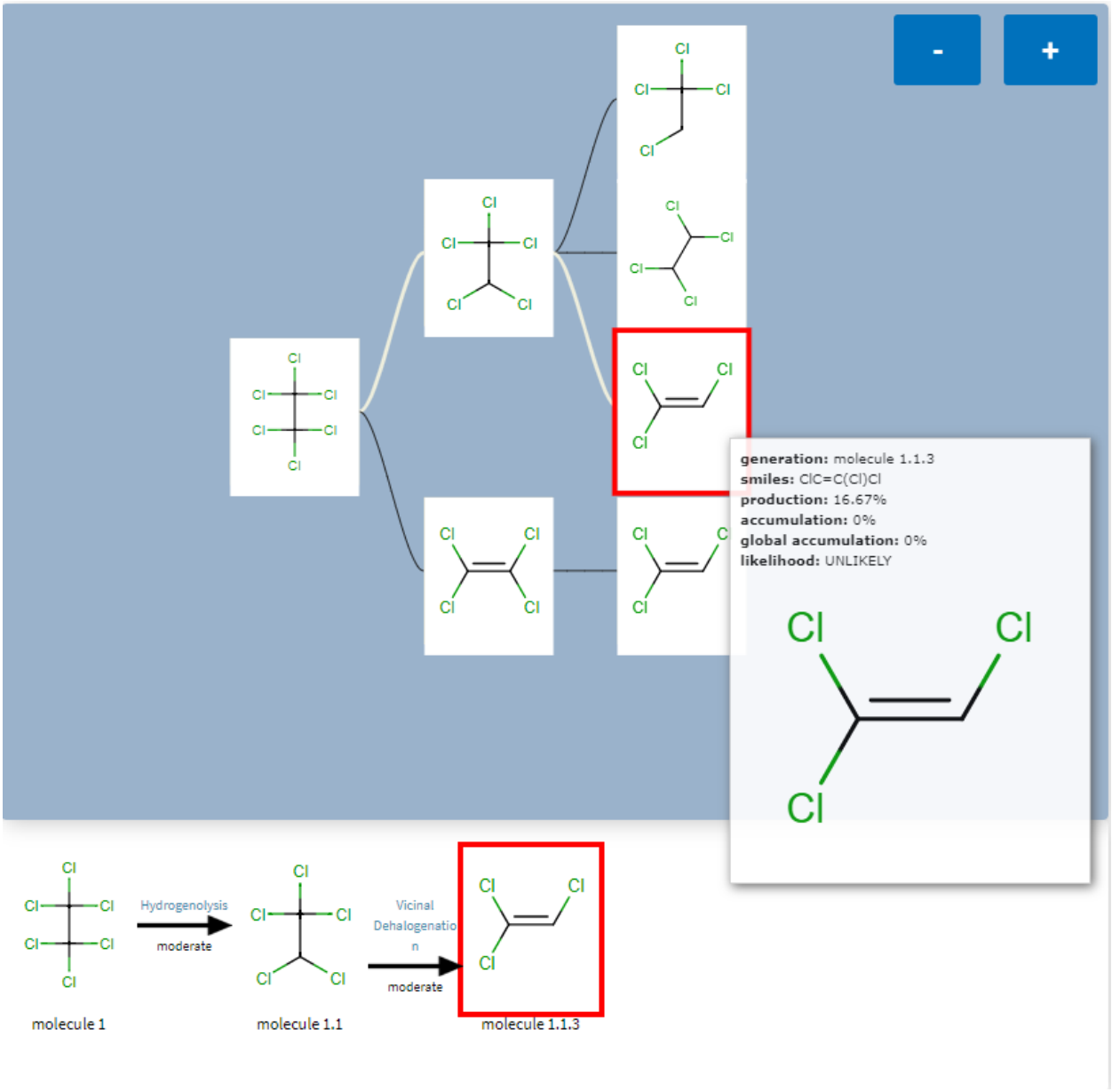


Regardless of tree type selected, all reaction pathways will initially display only the first generation of transformation products (the default value). The number of viewed generations can be increased by changing the number of generations in the "Display up to" window shown in the screen shot above. The screen shot below on the left illustrates the reaction pathway map for the formation of one generation of products of hexachloroethane based on execution of the abiotic hydrolysis and reduction libraries, as previously selected. The screen shot below on the right illustrates the reaction pathway map for the formation of two generation of products. Note that the number of observed generations cannot exceed the Generation Limit set on the previous screen.



By placing the cursor over a product, a popup box appears, with a molecule number that signifies its place in the reaction pathway map. For this example, tetrachloroethane (1.1.2) is the 2nd product formed in the second generation from the 1st product (i.e., pentachloroethane, 1.1), which was formed in the first generation from hexachloroethane as shown in the screen shot below. Below the molecule number are values for production, accumulation, global accumulation, and likelihood. The production, accumulation, and global accumulation values are explained in detail in a separate document available on the Reaction Library documentation page. Metabolites considered "likely" (having a global accumulation of at least 10%) will be highlighted with a blue border in the tree.

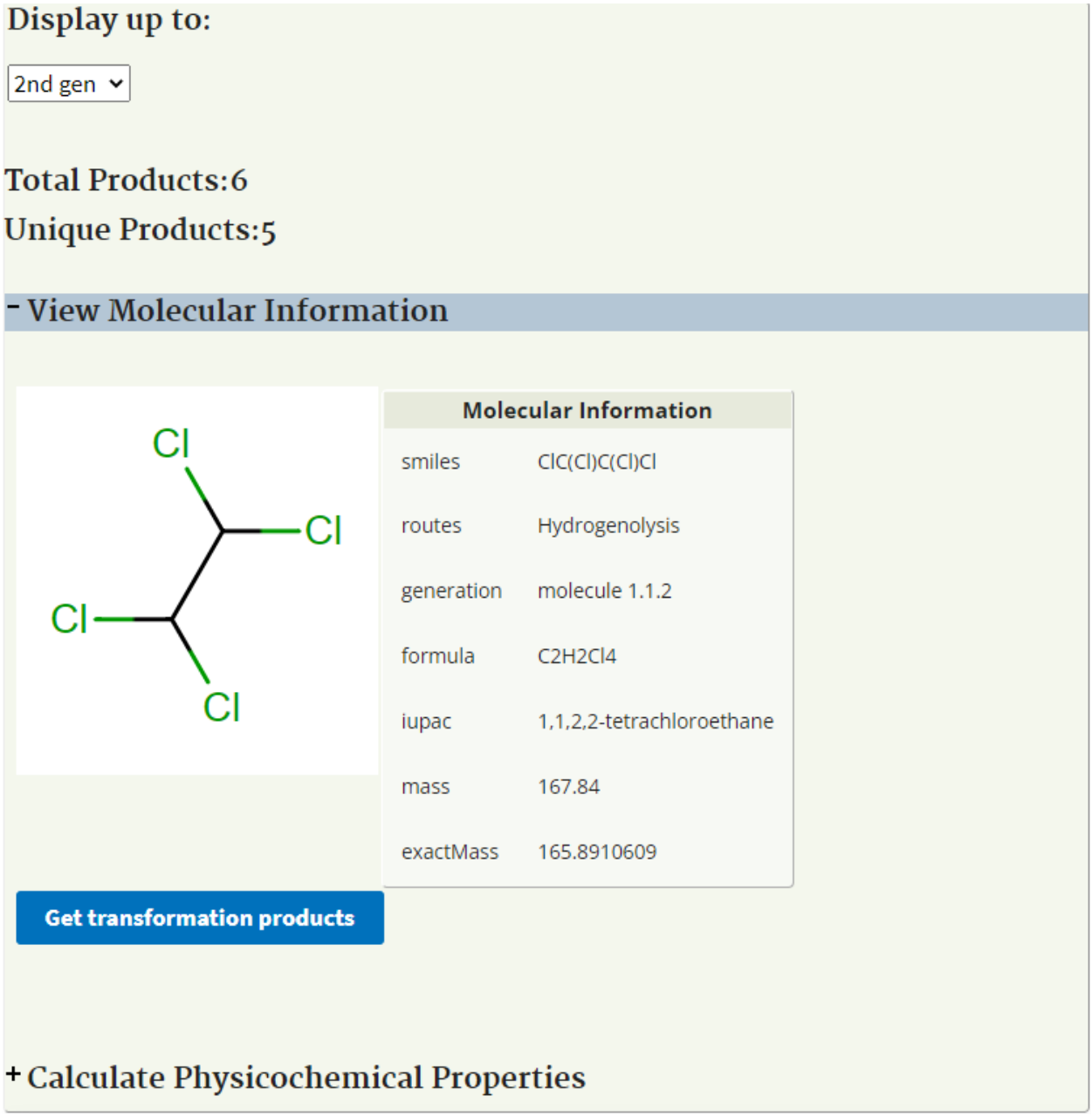
By left clicking on a product in the reaction pathway map, the next generation of transformation products that are predicted to form from a selected product, as well as the reactions that form them, are displayed under the reaction pathway map, as shown in the screen shot below. For each step in the transformation sequence, the name of the reaction scheme that generated the product is provided above the arrow between the parent and product. Clicking on a transformation scheme name opens a new browser tab with detailed information about the scheme from the Reaction Library documentation.



A qualitative description of the expected reaction rate for the reaction scheme is provided below the reaction arrow. This reaction rate designation is based on the median observed half-life for transformation of example molecules by the reaction scheme, as is shown in the table below. There can be considerable variability in the observed transformation rates for a given reaction scheme, so the actual transformation rate of the molecule of interest may be somewhat faster or slower than the median observed rate. For select schemes in the Abiotic Hydrolysis library, the user can select an option to replace the qualitative description with an estimated half-life value obtained from the EPI Suite [TM] HYDROWIN module (see section 6.6.1, page 42).

Scheme Rank	Half-life Range	Rate Description
7	$t_{½} < 30$ min	very fast
6	30 min $\leq t_{½} < 200$ min	fast
5	200 min $\leq t_{½} < 24$ hours	moderate
4	24 hours $\leq t_{½} < 7$ days	moderate
3	7 days $\leq t_{½} < 60$ days	slow
2	60 days $\leq t_{½} < 1$ year	slow
1	> 1 year	very slow

By right-clicking on a product, the molecular and metabolite information for the product is displayed above the reaction pathway map. The selected metabolite will be highlighted with a red border in the tree, as shown in the screenshot above. The “Get transformation products” button immediately below the molecular information box can be clicked to open a new browser tab with the product entered as the chemical of interest in the Generate Transformation Products workflow.



Below the Molecular Information table, clicking on the Calculate Physicochemical Properties dropdown box provides the various options for physicochemical properties and calculators to be applied to the selected transformation product, as shown in the screen shots below. The selected physicochemical properties will be calculated and displayed in the results table. For example, selection of the All and ChemAxon, EPI Suite™, TEST and OPERA buttons and Clicking on the Calculate Physicochemical Properties link provides the screen on the results for the selected physicochemical calculators for the selected metabolite, tetrachloroethane, as shown in the screen shot below.

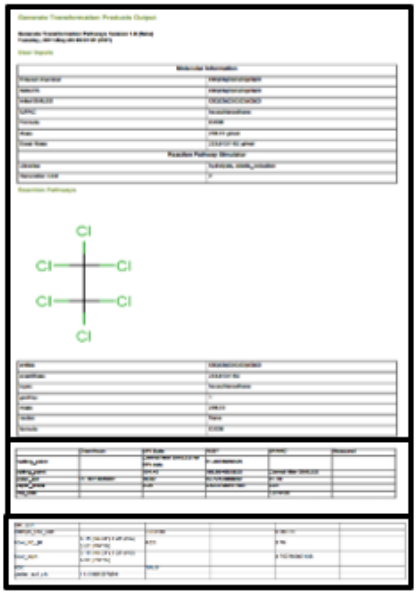

Calculate Physicochemical Properties
Calculate physicochemical properties for: selected metabolite
Select physicochemical properties to gather for selected metabolite, then click "Calculate data" below..
ChemAxon
EPI Suite
TEST
OPERA
Geometric Mean
Measured
All
Neutral Species Properties:
Melting Point (°C)
-48.66
-48.80 HC -44.60 NN -27.90 GC
-39.83
-42.65
-43.80
Boiling Point (°C)
149.90
140.60 HC 115.90 NN 140.50 GC
147.58
136.53
146.50
Water Solubility (mg/L)
4.29e+2
1.25e+3 WSKOW 1.85e+3 WATERNT
2.73e+3 HC 2.92e+3 NN 7.42e+2 GC
2.96e+3
1.34e+3
2.83e+3
Vapor Pressure (mmHg)
4.72e+0
5.06e+0 HC 6.01e+0 NN 5.13e+0 GC
7.29e+0
5.21e+0
1.33e+1
Ionization Constant
none
none
Henry's Law Constant (atm-m3/mol)
1.51e-3
5.01e-4
1.51e-3
3.67e-4
Octanol/Water Partition Coefficient (log)
3.54 KLOP 2.06 VG 2.91 PHYS
2.19
2.35
2.68
2.39
Organic Carbon Partition Coefficient (log(L/kg))
1.98 MCI 2.07 KOW
1.89
2.03
Bioconcentration Factor (log(L/kg))
1.24 REG 1.29 A-G
1.36 SM 1.04 HC 1.09 NN 1.27 GC
1.07
1.22
Bioaccumulation Factor (log(L/kg))
1.29 A-G
1.29
pH-Dependent Properties at pH: 7.4
Octanol/Water Distribution Coefficient (log)
3.54 KLOP 2.06 VG 2.91 PHYS
2.35
2.84
Water Solubility (mg/L)
4.29e+2
4.29e+2
Available
Unavailable
Clear data
Calculate data

To get physicochemical data for multiple metabolites, select the option for calculating physicochemical properties for up to the first, second, or third generation of metabolites, or for all calculated metabolites from the drop-down menu. Then, select the properties and calculators to be used and click the Calculate data button. The results for multiple metabolites will not be presented in the table as they are for a single metabolite. To view the results for multiple metabolites, download and view the report (as a PDF, CSV, or HTML file) as described below.

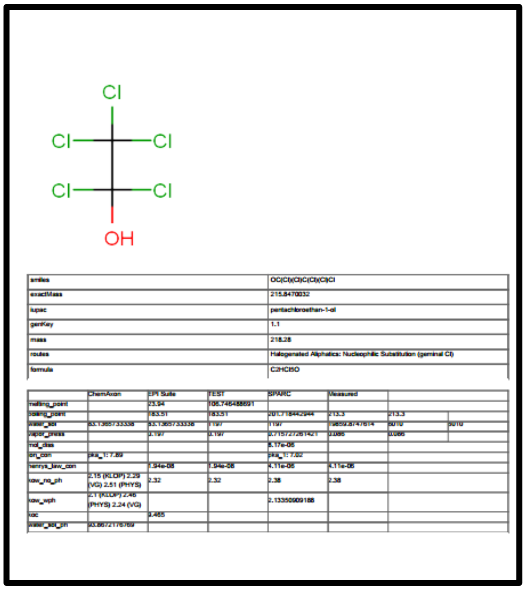
6.7. Generation of PDF, HTML and CSV Reports

The .pdf, .html and .csv buttons appear on the top right corner of the results page, regardless of the workflow. Clicking on the .pdf button generates a PDF file that can be viewed in the web browser or using free PDF software. The HTML file can be viewed using a web browser.

The PDF and HTML reports are multi-page reports showing the calculated physicochemical data for the parent compound and the selected transformation products. Examples of the PDF and HTML reports are shown below in the first two screen shots.



PDF Report



smiles	OC(Cl)(Cl)C(Cl)(Cl)Cl
exactMass	215.8470032
iupac	pentachloroethan-1-ol
genKey	1.1
mass	218.28
routes	Halogenated Aliphatics: Nucleophilic Substitution (geminal Cl)
formula	C2HCl5O

	ChemAxon	EPI Suite	TEST	SPARC	Measured		
melting_point		23.94	106.746488691				
boiling_point		183.51	183.51	201.718442944	213.3	213.3	
water_sol	53.1365733336	53.1365733336	1197	1197	19659.8747614	5010	5010
vapor_press		0.197	0.197	0.715727261421	0.066	0.066	
mol_diss				5.17e-06			
ion_con	pka_1: 7.89			pka_1: 7.02			
henrys_law_con		1.94e-08	1.94e-08	4.11e-06	4.11e-06		
kow_no_ph	2.15 (KLOP) 2.29 (VG) 2.51 (PHYS)	2.32	2.32	2.38	2.38		
kow_wph	2.1 (KLOP) 2.46 (PHYS) 2.24 (VG)			2.13350909188			
koc		9.465					
water_sol_ph	93.8672176769						

HTML Report

The CSV report is generated in a tabular format as shown below in the screen shot.

	A	B	C	D	E	F	G
1	genKey	routes	smiles	iupac	formula	mass	exactMass
2	1		ClC(Cl)(Cl)	hexachlor	C2Cl6	236.72	233.8131
3	1.1	Halogenat	OC(Cl)(Cl)	pentachlo	C2HCl5O	218.28	215.847
4	1.2	Hydrogen	ClC(Cl)C(C	1,1,1,2,2-p	C2HCl5	202.28	199.8521
5	1.2.1	Hydrogen	ClC=C(Cl)(	1,1,2-trich	C2HCl3	131.38	129.9144
6	1.2.2	Hydrogen	ClC(Cl)=C(	tetrachlor	C2Cl4	165.82	163.8754
7	1.2.3	Hydrogen	OC(Cl)(Cl)	1,1,2,2-tet	C2H2Cl4O	183.84	181.886
8	1.2.4	Hydrogen	OC(Cl)C(C	1,2,2,2-tet	C2H2Cl4O	183.84	181.886
9	1.2.5	Hydrogen	ClCC(Cl)(C	1,1,1,2-tet	C2H2Cl4	167.84	165.8911
10	1.2.6	Hydrogen	ClC(Cl)C(C	1,1,2,2-tet	C2H2Cl4	167.84	165.8911
11	1.3	Vicinal De	ClC(Cl)=C(	tetrachlor	C2Cl4	165.82	163.8754
12	1.3.1	Vicinal De	ClC=C(Cl)(	1,2-trich	C2HCl3	131.38	129.9144

CSV Report

7. References

Drossman, H.; Johnson, H.; Mill, T. 1988. Structure activity relationships for environmental processes 1: Hydrolysis of esters and carbamates. *Chemosphere* 17(8): 1509-1530.

Hou, T.J., Xia, K., Zhang, W., Xu, X.J. 2004. ADME evaluation in drug discovery. 4. prediction of aqueous solubility based on atom contribution approach, *J. Chem. Inf. Comput. Sci.* 44: 266-275.

Klopman, G., Li, J.-Y., Wang, S., Dimayuga., M. 1994. Computer automated log P calculations based on an extended group contribution approach. *J. Chem. Inf. Comput. Sci.* 34: 752-781.

Mansouri, K., Grulke, C.M., Judson, R.S., Williams, A.J. 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. *J Cheminform.* 10:10.

Martin, T.M., Harten, P., Venkatapathy, R., Das, S. Young, D.M. 2008. A hierarchical clustering methodology for the estimation of toxicity, toxicology mechanisms and methods. *Toxicol. Mech. Methods* 18: 251–266.

Martin, T.M., Young. D.M. 2001. Prediction of the acute toxicity (96-h LC50) of organic compounds in the fathead minnow (Pimephales Promelas) using a group contribution method. *Chem. Res. Toxicol.* 14: 1378–1385.

Meylan, W., P.H. Howard and R.S. Boethling. 1992. Molecular topology/fragment contribution method for predicting soil sorption coefficients. *Environ. Sci. Technol.* 26: 1560-1567.

Meylan W.M., Howard, P.H. 1995. Atom/Fragment contribution method for estimating octanol-water partition coefficients, *J. Pharm. Sci.*, 84: 83-92.

Meylan, W.M., Howard, P.H., Boethling, R.S. 1996. Improved method for estimating water solubility from octanol/water partition coefficient. *Environ. Toxicol. Chem.* 15: 100-106.

Szegezdi, J., Csizmadia, F. 2004. Prediction of dissociation constant using microconstants. 27th ACS National Meeting, Anaheim, California, March 28-April 1, 2004.

Szegezdi, J., Csizmadia, F. 2007. A method for calculating the pKa values of small and large molecules. American Chemical Society Spring meeting, March 25-29th, 2007.

US EPA. 2012. Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. United States Environmental Protection Agency, Washington, DC, USA.

U.S. EPA 2020. User's Guide for T.E.S.T. (version 5.1) (Toxicity Estimation Software Tool): A Program to Estimate Toxicity from Molecular Structure, EPA/600/R-16/058. United States Environmental Protection Agency, Washington, DC, USA.

Viswanadhan, V.N., Ghose, A.K., Revankar, G.R., Robins, R.K. 1989. Atomic physicochemical parameters for three-dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* 29: 163-172.