User's Guide for the Chemical Transformation Simulator (CTS) (β-version 1.0)

5/10/17

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

2017 U.S. Environmental Protection Agency

The Chemical Transformation Simulator (CTS) User's Guide is designed to provide the first time user a complete understanding of how to use 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

Introduction 2

Background 3

Using the CTS Software 4

- Accessing the CTS 4
- Entering Chemical in Single or Batch Mode 7
- Single Chemical Entry 8
- Batch Chemical Entry 9
- Execution of the CTS Workflows 10
  - Calculate Chemical Speciation Workflow 10
- Calculate Ionization Constants 13
- Calculate Dominant Tautomer Distribution 17
- Calculate Stereoisomers 20
- Calculate p-Chem Properties Workflow 25
- Generate Transformation Products Workflow 31
- Reaction System Guidelines 34
- Generation of .pdf , .html and .csv Reports 40

Introduction

The Chemical Transformation Simulator (CTS) provides the calculated physico-chemical 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, computational chemistry tools for the calculation of physico-chemical properties, and software technologies that provide access to on-line databases for environmental descriptors required for estimating environmental concentrations.

The β-version 1.0 of the CTS consists of 3 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. Additionally, the Earth Systems Model (ESM) and Reaction Rate Calculator (RRC), will be functional in the deployable version of the CTS.

β-Version 1.0

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

Deployable Version

- Structure-based Database (SBD): Populated with calculated and measured physico-chemical properties of parent and potential transformation products
- Earth Systems Model (ESM): Provides data mining abilities for environmental descriptors such as pH and temperature
- Reaction Rate Calculator (RRC): Calculates transformation products based on the parameterization and execution of QSARs and Algorithms

Background

A key Agency need identified as a high priority in the Chemical Sustainability and Safety (CSS) research program is for high throughput computational systems to simulate environmental fate and transport for organic chemicals for which environmental 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 development of the physico-chemical properties calculator (PPC) and the Reaction Pathway Simulator (RPS). The PPC is based on a consensus approach that would allow the user to compare output generated by a number of calculators that take different approaches to calculating specific physicochemical properties. The calculators we are currently accessing include (1) SPARC (SPARC Performs Automated Reasoning in Chemistry), which uses a mechanistic-based approach, (2) EPI Suite, which uses a fragment-

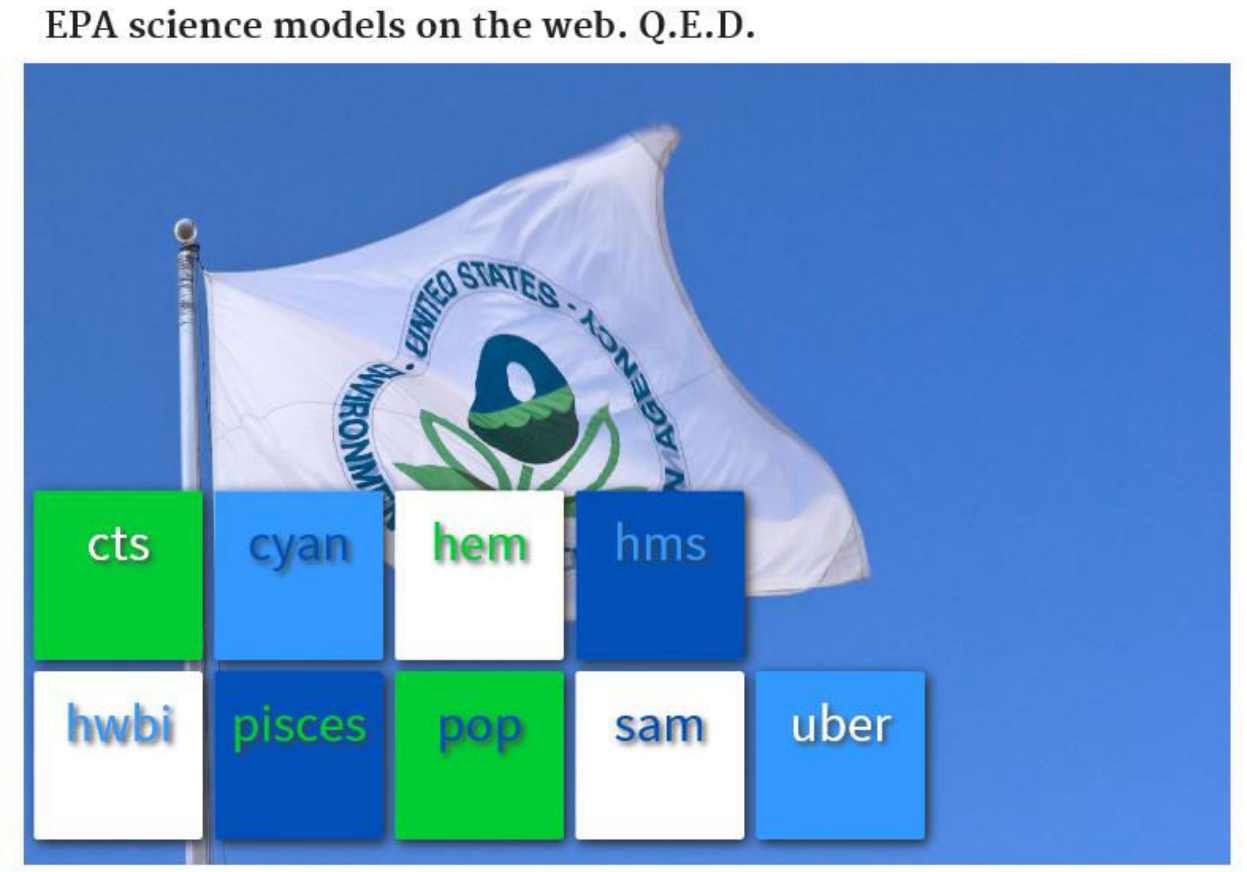
based approach, (3) TEST (Toxicity Estimation Software Tool), which uses QSAR-based approaches, and (4) ChemAxon plug-in calculators, which use an atom-based fragment approach. The output derived from these calculators will enable the user to compare the calculated data with measured data in readily 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 of reactive functional groups (i.e., reduction and hydrolysis). These one-step reactions represent viable transformation pathways based on the identification and subsequent transformation of reactive functional groups. A reaction library for human metabolism for phase 1 transformations developed by ChemAxon is also available through the CTS. The development of reaction libraries allow us to "encode" the known process science published (current and future) in the peer-reviewed literature. The encoding of process science is accomplished through the use of Chemical Terms Language and Smart Reaction Smile string 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.

Using the CTS Software

Accessing the CTS

Currently only EPA intranet users have access to the CTS. The CTS can be accessed through https://qedinternal.epa.gov, which provides the models currently available on the web Q.E.D.



CTS Homepage

The home page of the CTS is accessed by clicking on the CTS box. The home page provides access to the CTS through the selection of one of three CTS workflows and general information concerning the major components of the CTS and the physicochemical calculators as shown below.

Links to the process science supporting the currently available reaction libraries are also available. For example, clicking on the Abiotic Hydrolysis tab, followed by clicking "Click here to download the abiotic hydrolysis library" provides the following screen:

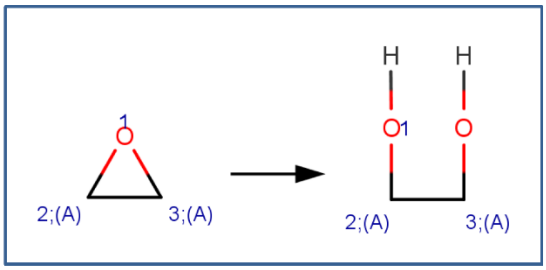
Abiotic Hydrolysis Reaction Library

Version 1.6 of the Abiotic Hydrolysis Reaction Library contains 25 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
- Amide Hydrolysis
- Lactam Hydrolysis
- Carbamate Hydrolysis
- Urea Hydrolysis
- Cyclic Urea Hydrolysis
- Sulfonylurea Hydrolysis
- Thiocarbamate Hydrolysis
- Nitrile Hydrolysis
- N-S Cleavage
- Imide Hydrolysis
- Acid Halide Hydrolysis
- Dehydration of Geminal Diols

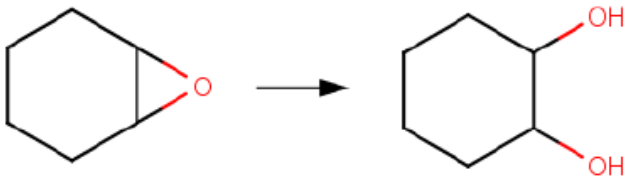
Selection of one of the transformation pathways provides the reaction scheme, and documented examples with references. The following illustrates this information for Epoxide Hydrolysis:

SCHEME:

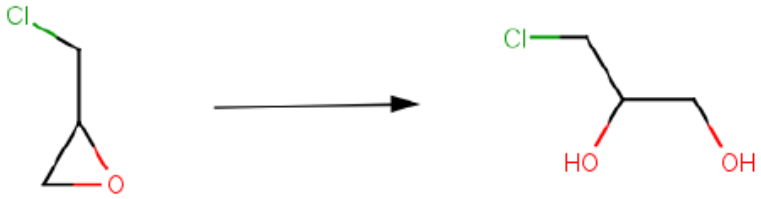


EXAMPLES:

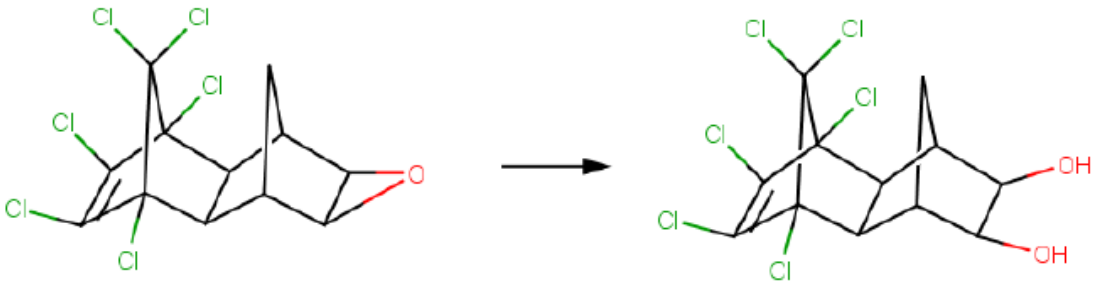
- 1,2-Epoxycyclohexane (McMurry, 2011)



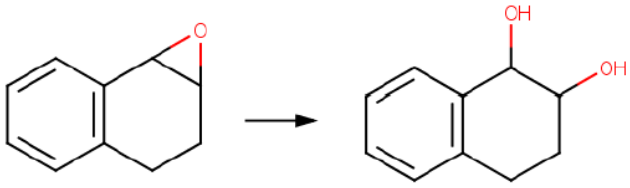
- Epichlorohydrin (Gaca *et al*, 2011)



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



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

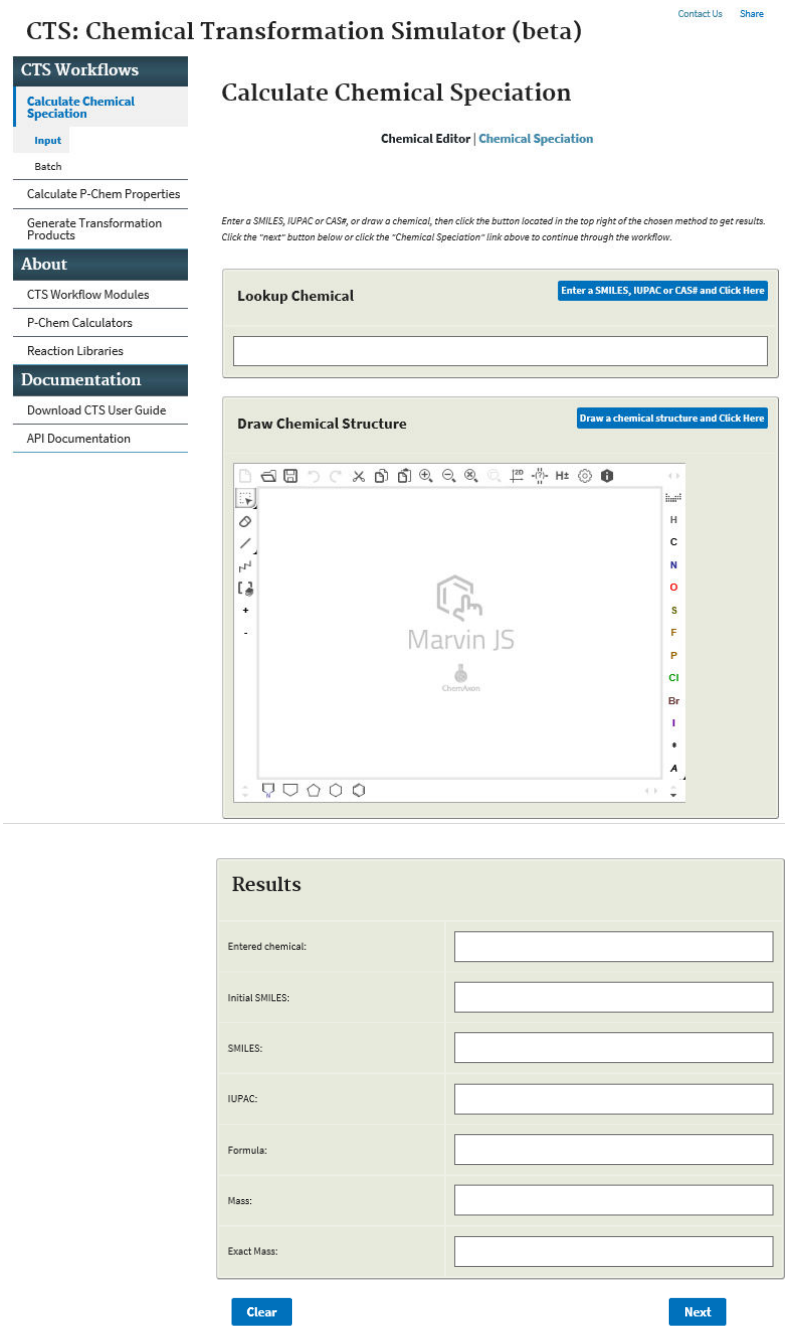


Entering Chemical in Single or Batch Mode

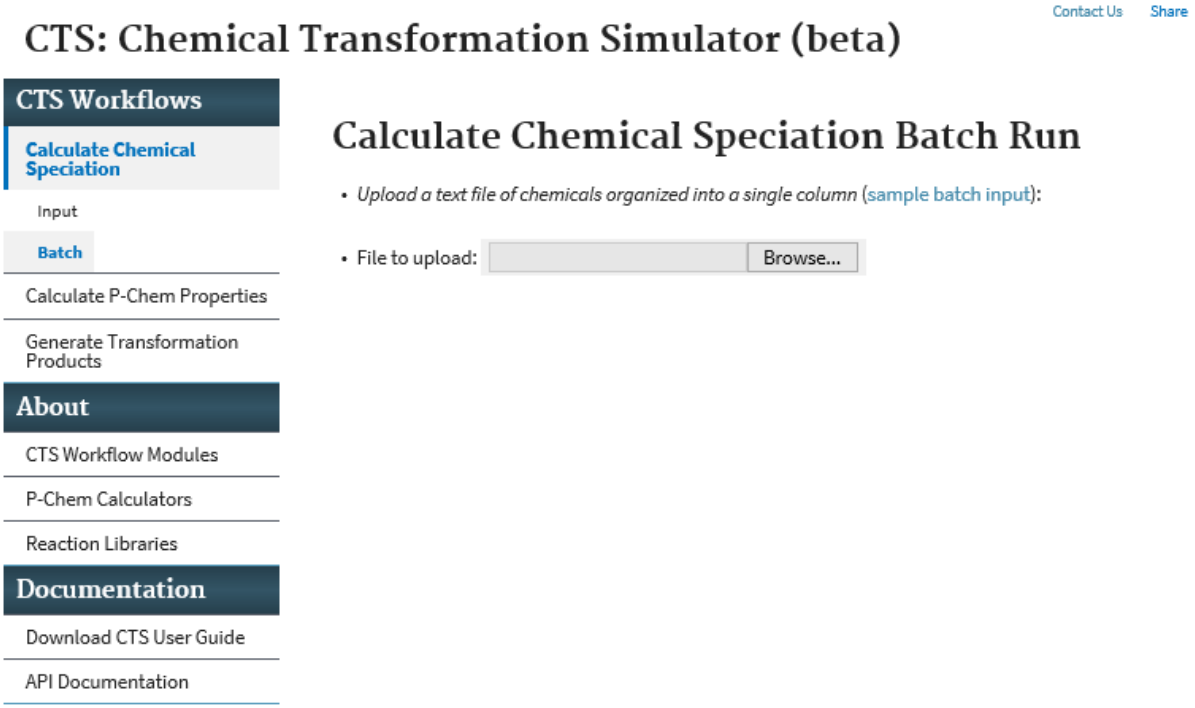
The CTS is executed by selecting one of three available workflows (see descriptions below) and by entering a single chemical or by batch mode. The process for entering a single chemical or by batch as described below is identical for each of the workflows.

Single Chemical Entry

For single chemical entry, the Input tab is selected at the top of the workflow frame. The Chemical Editor appears where there is the option to either enter a SMILES String, IUPAC chemical name, or CAS# in the Lookup Chemical box, or to draw a chemical structure using the Chemical Editor (see below). For either case, the appropriate box must be clicked after providing the required information. Details concerning the use of the chemical editor can be found at https://docs.chemaxon.com/display/marvinsketch/MarvinSketch+User%27s+Guide.



Batch Chemical Entry



By clicking on the sample batch input text, the example batch file is illustrated below. The chemicals are entered into a single column of SMILES strings. The default value for the β-version is currently set to a maximum of 10 chemicals. This default value will be increased once testing of the β-version has been completed.

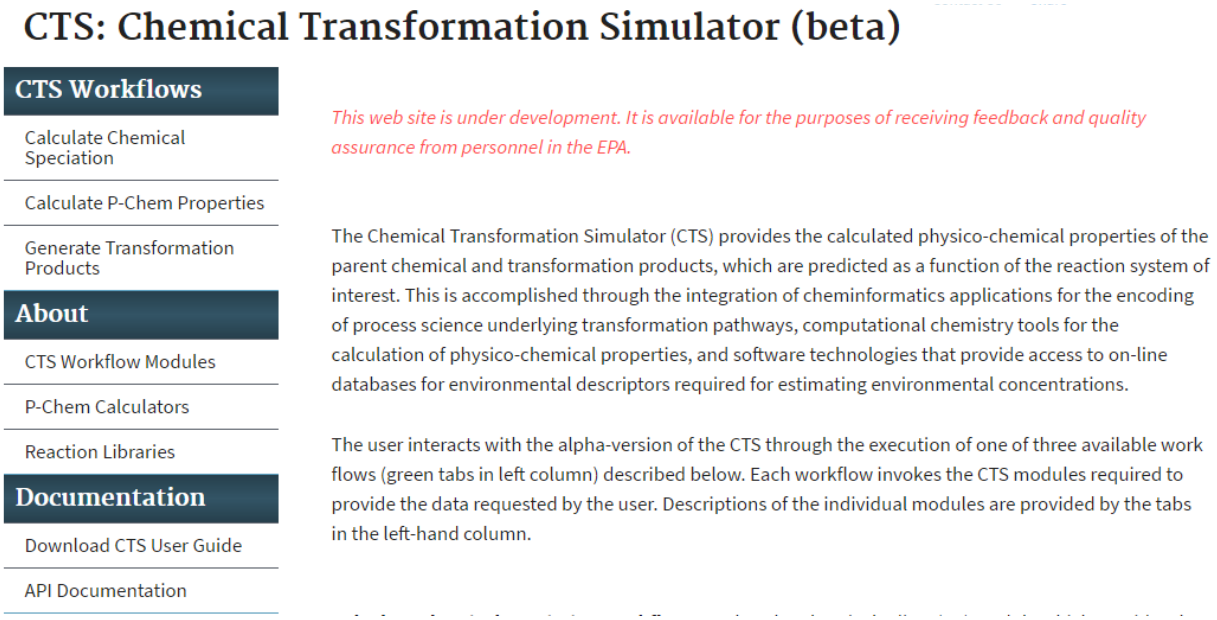
Sample batch input:

```
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
```

Execution of the CTS Workflows

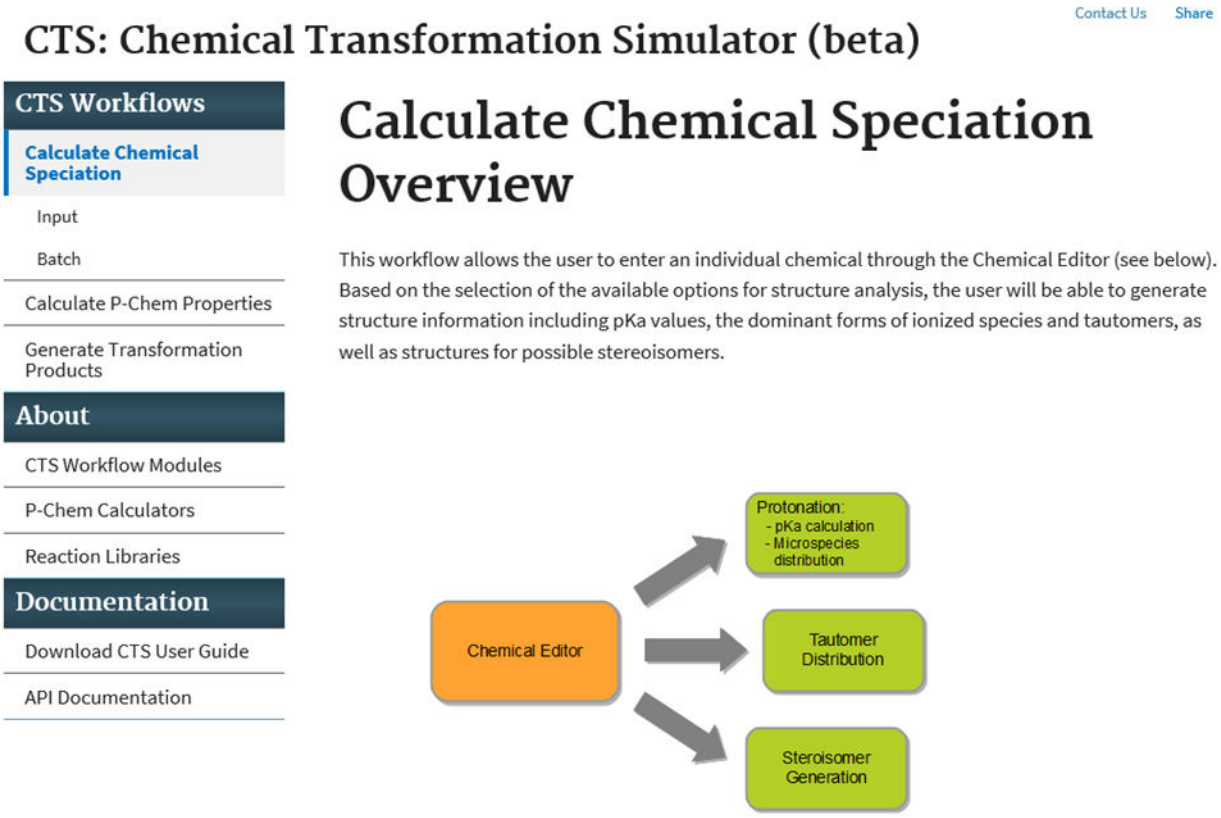
The user executes the CTS through the selection of one of three available workflows:

- Calculate Chemical Speciation
- Calculate p-Chem Properties
- Generate Transformation Products

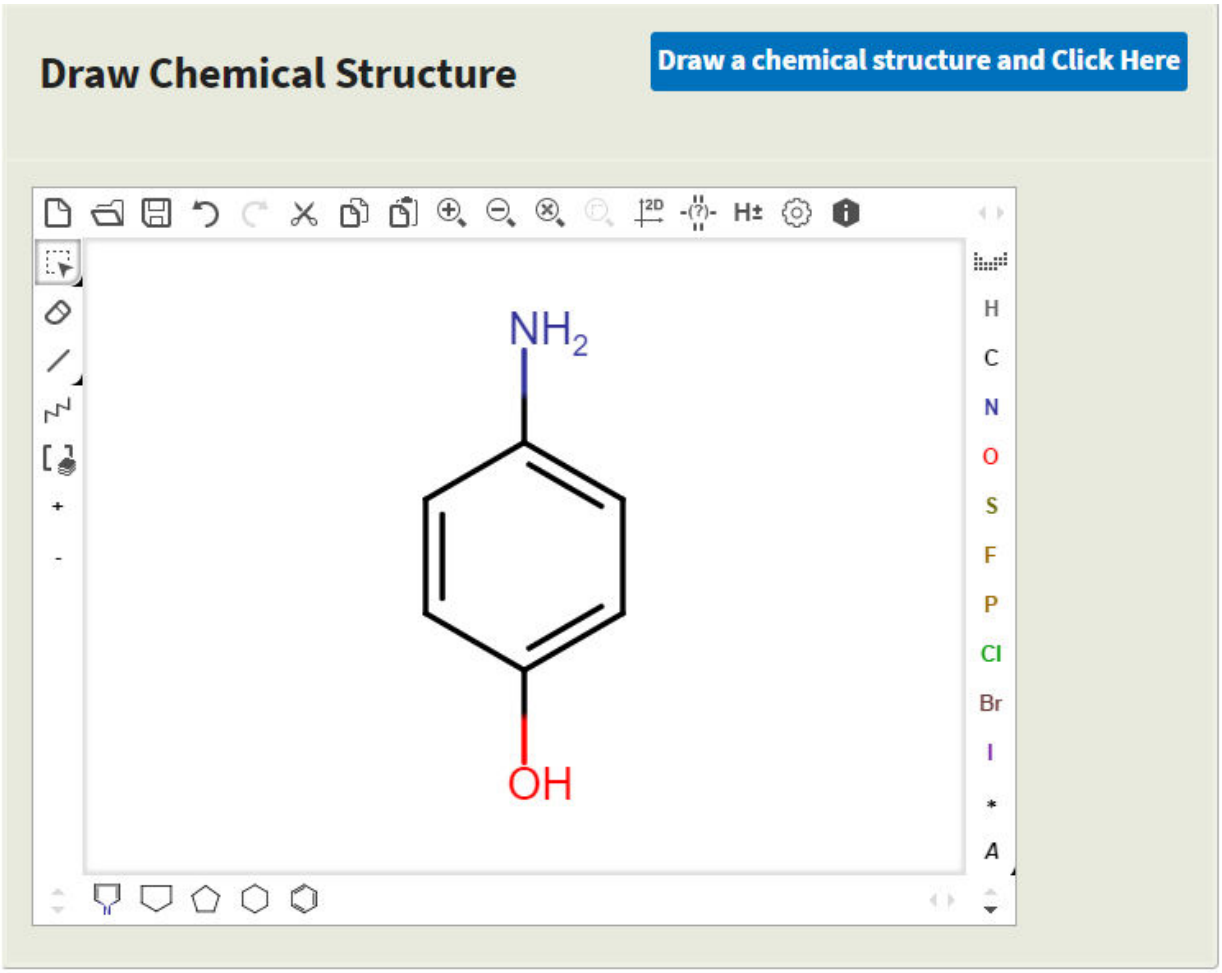


Calculate Chemical Speciation Workflow

Selection of the Calculate Chemical Speciation Workflow provides this page illustrating the workflow overview as illustrated below.



Clicking on Single Chemical Entry button, the user is taken to the chemical editor. For the following example, 4-aminophenol was entered into the Chemical Editor.



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:

- Calculate Ionization Constants
- Calculate Dominant Tautomer Distribution
- Calculate Stereoisomers

Select any combination of the calculators; use the provided default values or to 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.

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 the purpose of this demonstration.

Calculate Chemical Speciation

Chemical Editor | **Chemical Speciation**

Check one or more calculation methods to run, then hit submit below

Calculate Ionization Constants (p*Ka*) Parameters	
Number of decimals for pKa:	2
pH Lower Limit:	0
pH Upper Limit:	14
pH Step Size:	0.2
Generate Major Microspecies at pH:	7.0
Isoelectric Point (pI) pH Step Size for Charge Distribution:	0.5

Calculate Dominant Tautomer Distribution	
Maximum Number of Structures:	100
at pH:	7.0

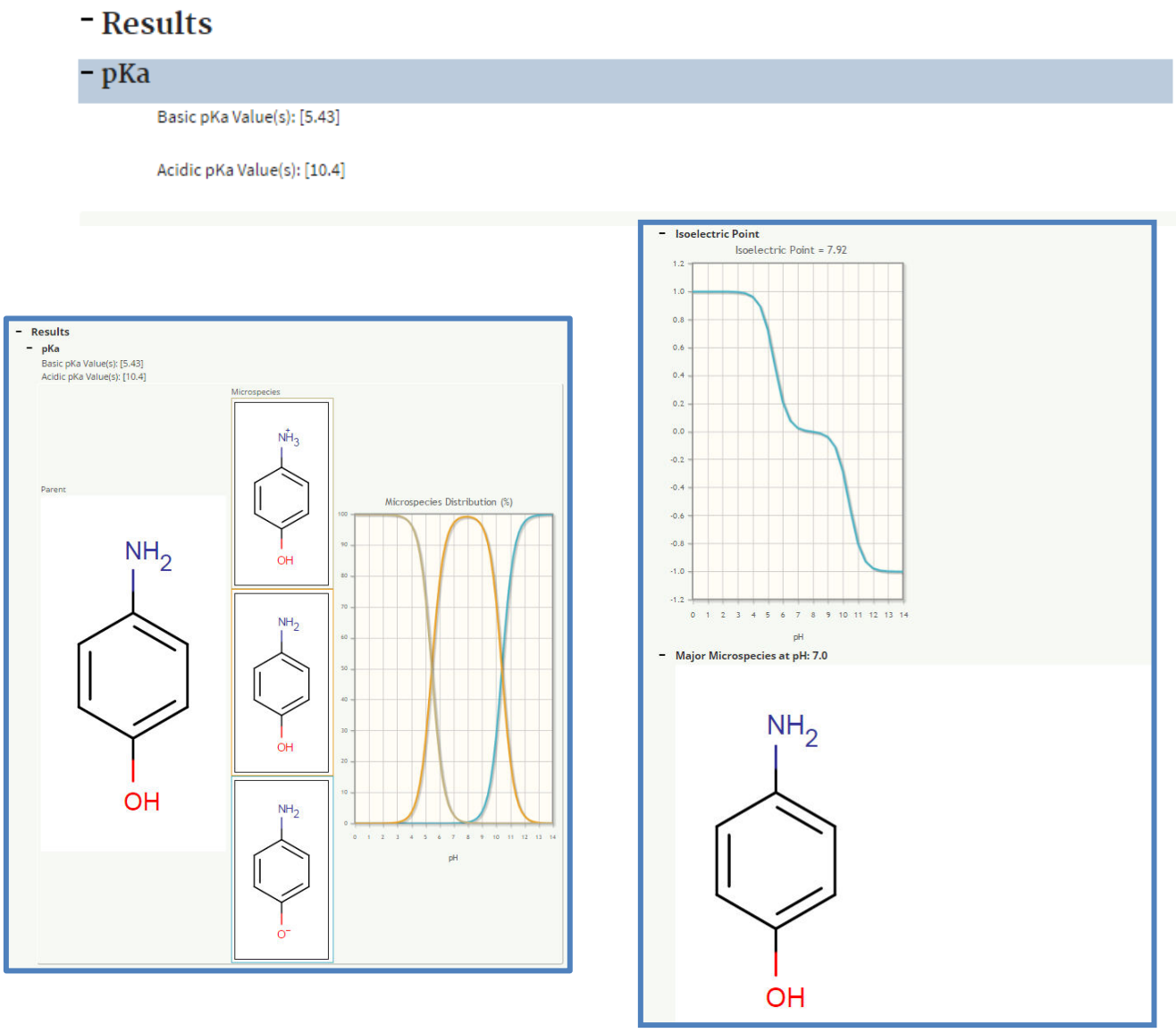
Calculate Stereoisomers	
Maximum Number of Structures:	100

Defaults Clear Back Submit

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

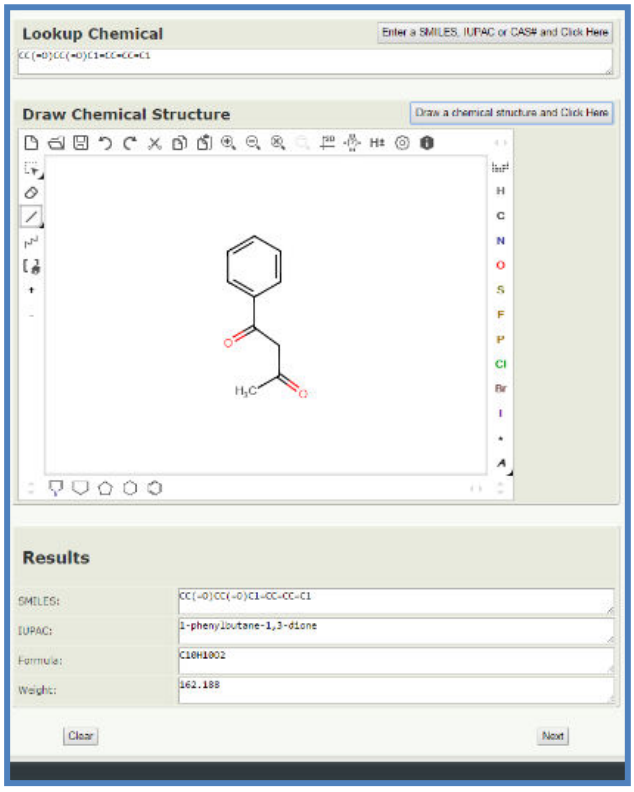
Inputs: The molecular information and ionization parameters provided by the user.

- **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.

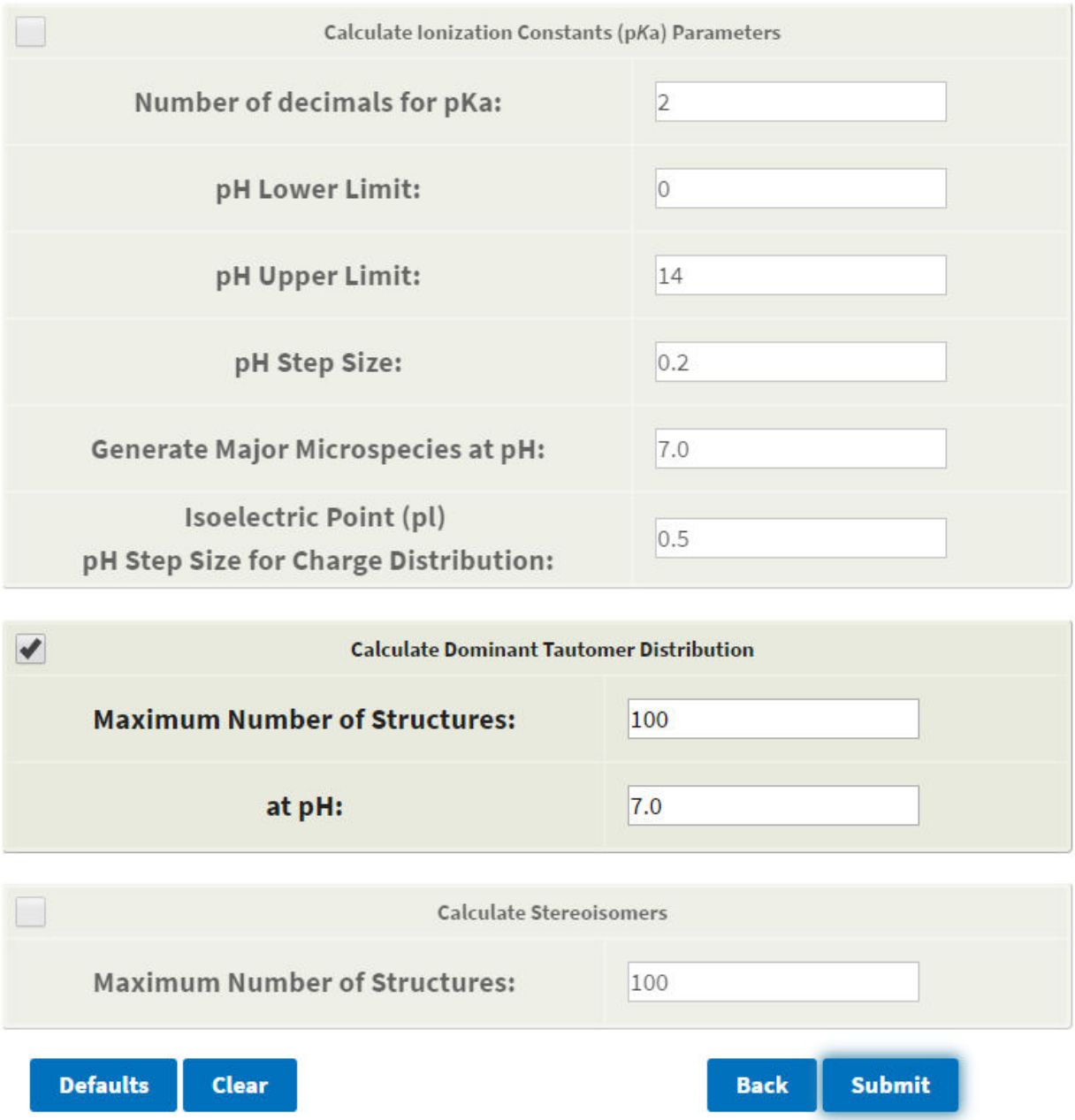


Calculate Dominant Tautomer Distribution

For the purpose of this demonstration, 1-phenylbutane-1,3-dione has been entered into the Chemical Editor.

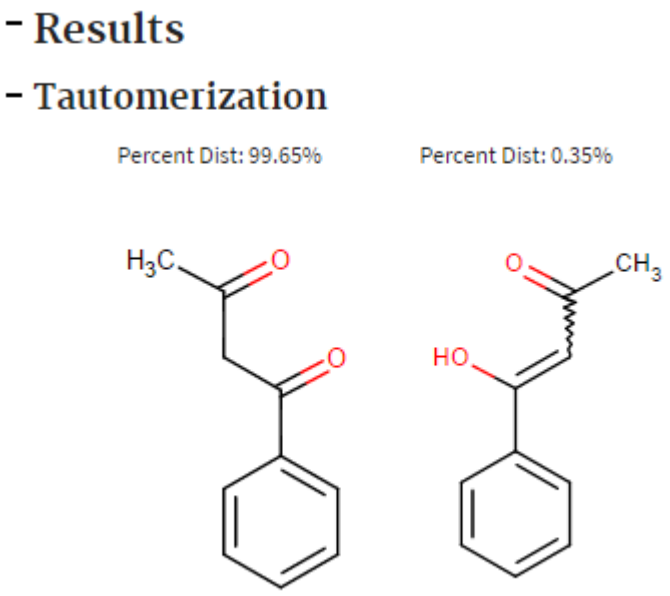


Clicking the Next button brings up the Calculate Chemical Speciation Workflow Inputs page. After selecting the Calculate Dominant Tautomer Distribution option, enter a limit for the number of possible tautomers and the pH value for which the distribution will be calculated. The default values are pH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



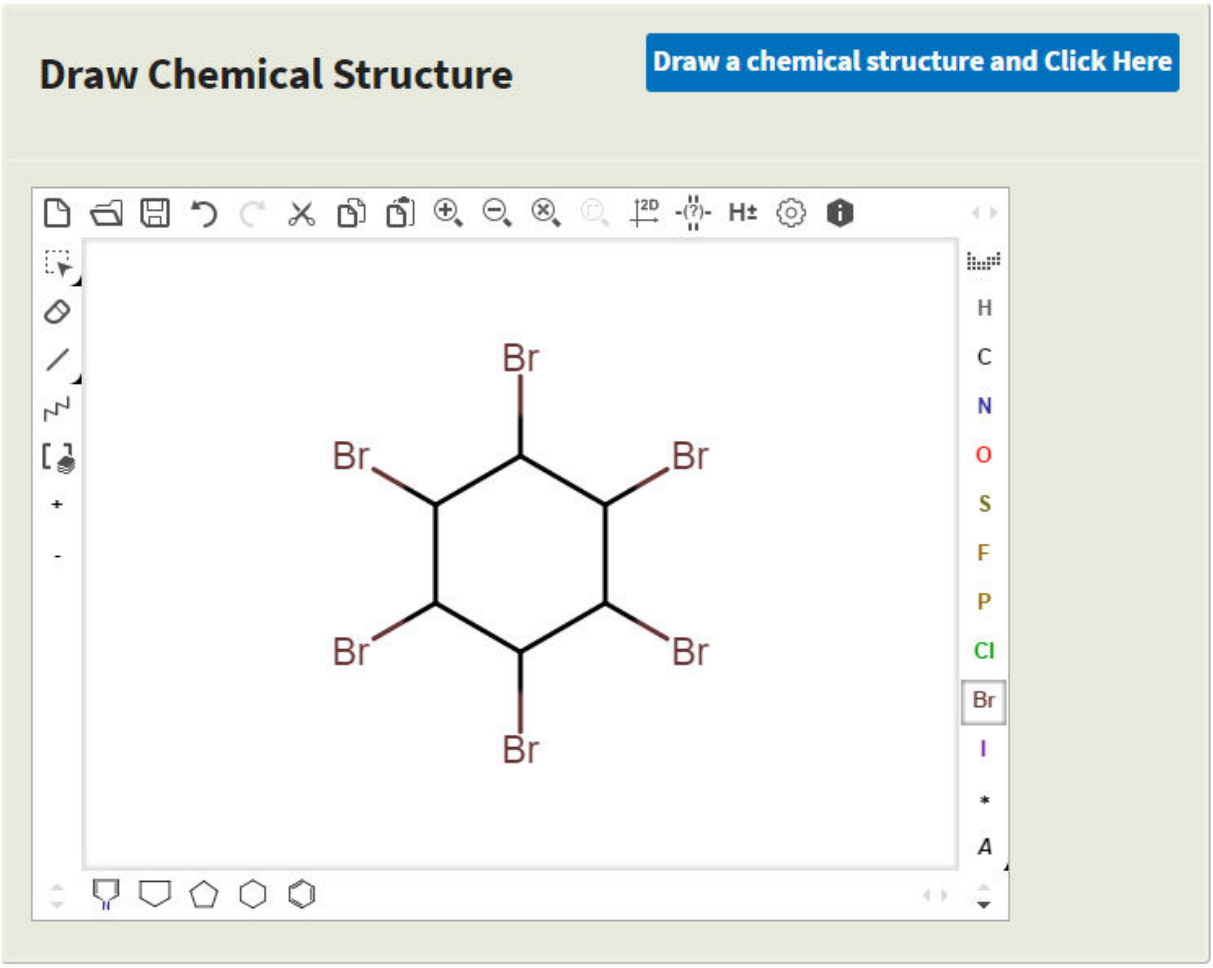
Click on the Submit button to view the output page for the tautomerization distribution based on the user-defined values.

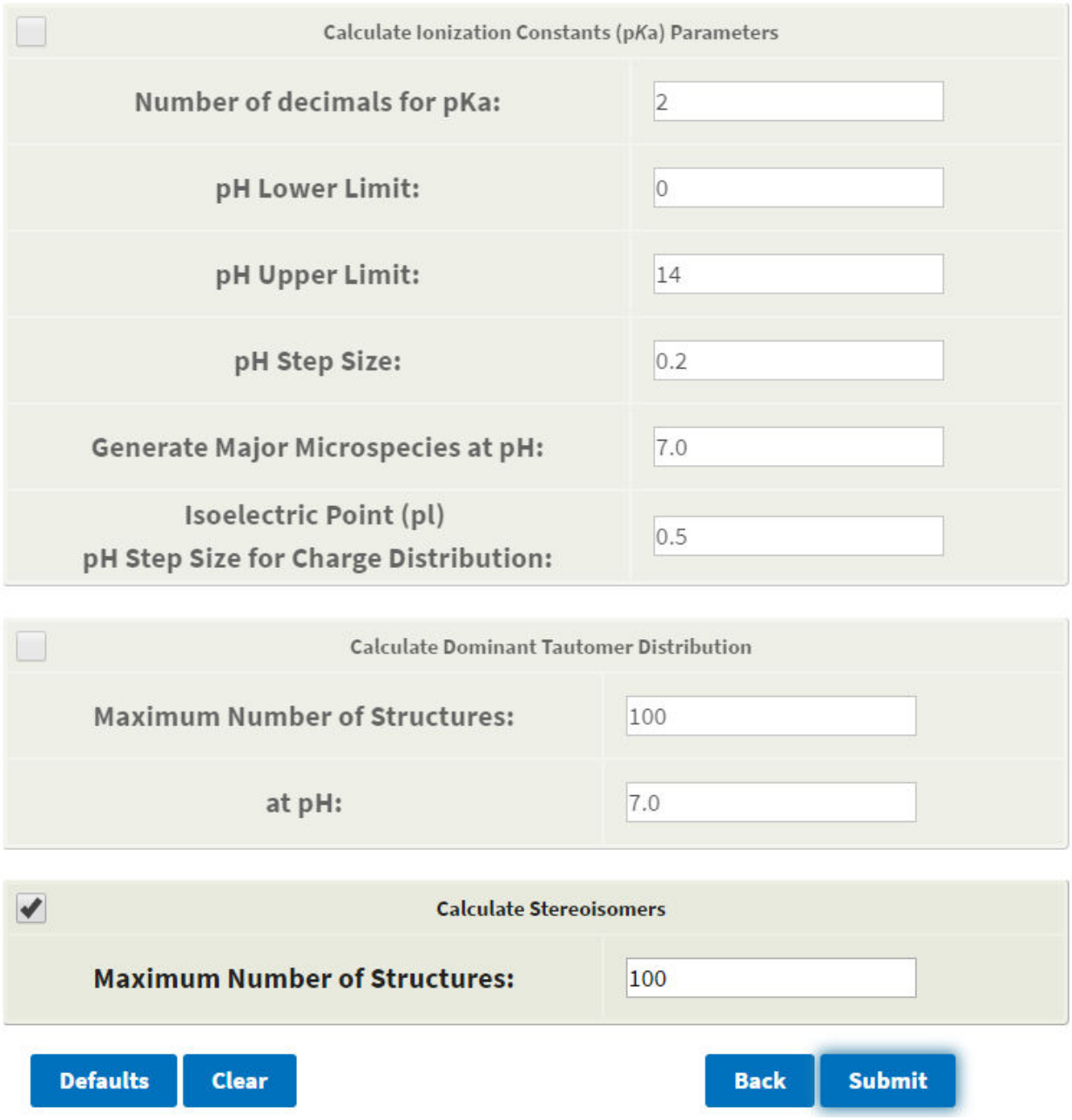
The Output screen shows the User Inputs (see above) as well as the tautomer distribution for the chemical of interest. The individual structures can be enlarged by placing the cursor on top of the structure. The molecular information including the formula, IUPAC name, mass and SMILES string is also provided.



Calculate Stereoisomers

For the purpose of this demonstration, 1,2,3,4,5,6-hexabromocyclohexane has been entered into the Chemical Editor.



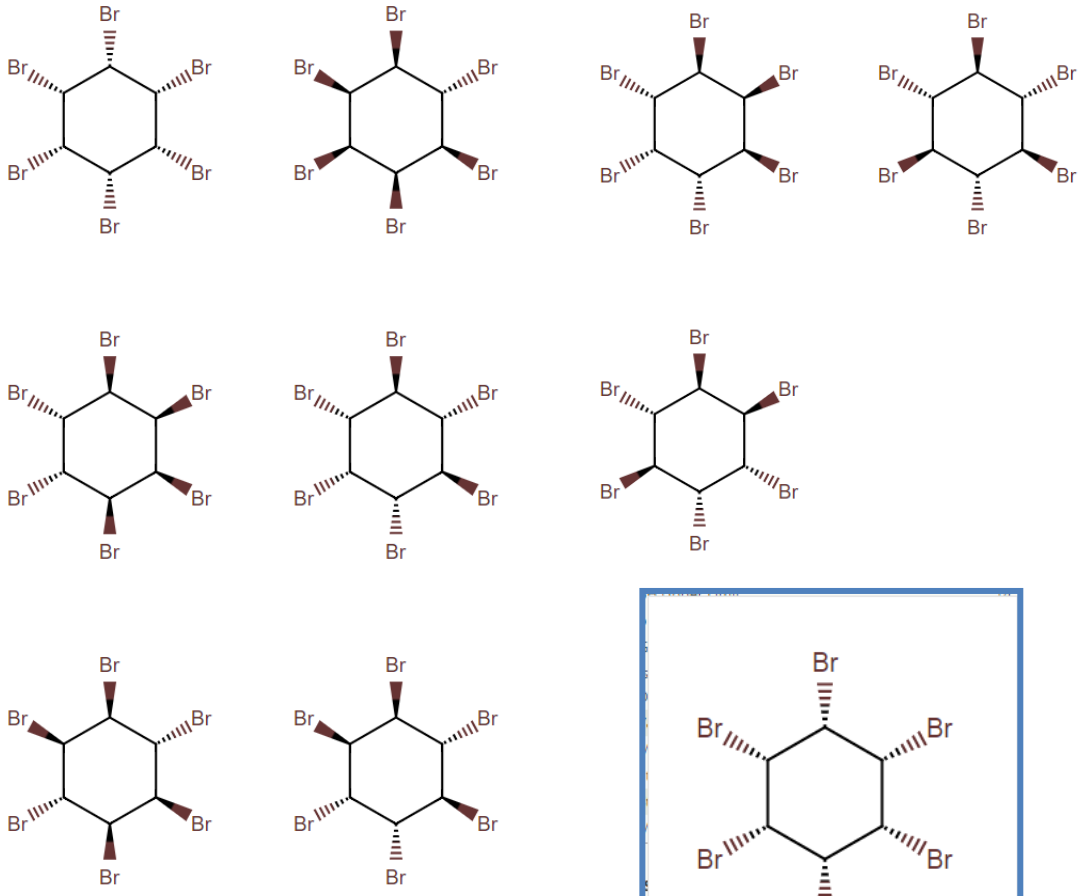


After selecting the Calculate Stereoisomers option, enter a limit for the maximum number of possible stereoisomers. The default value is 100 stereoisomers.

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 isomers. 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.

Results

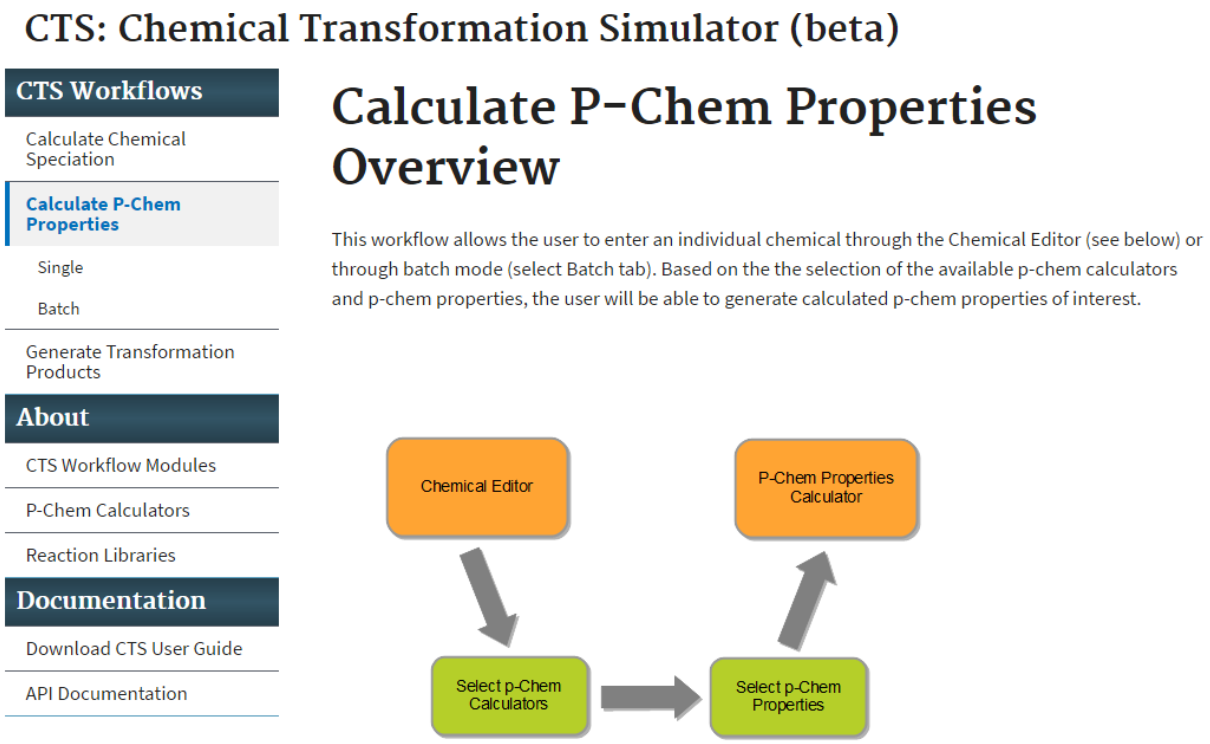
Stereoisomers (9)



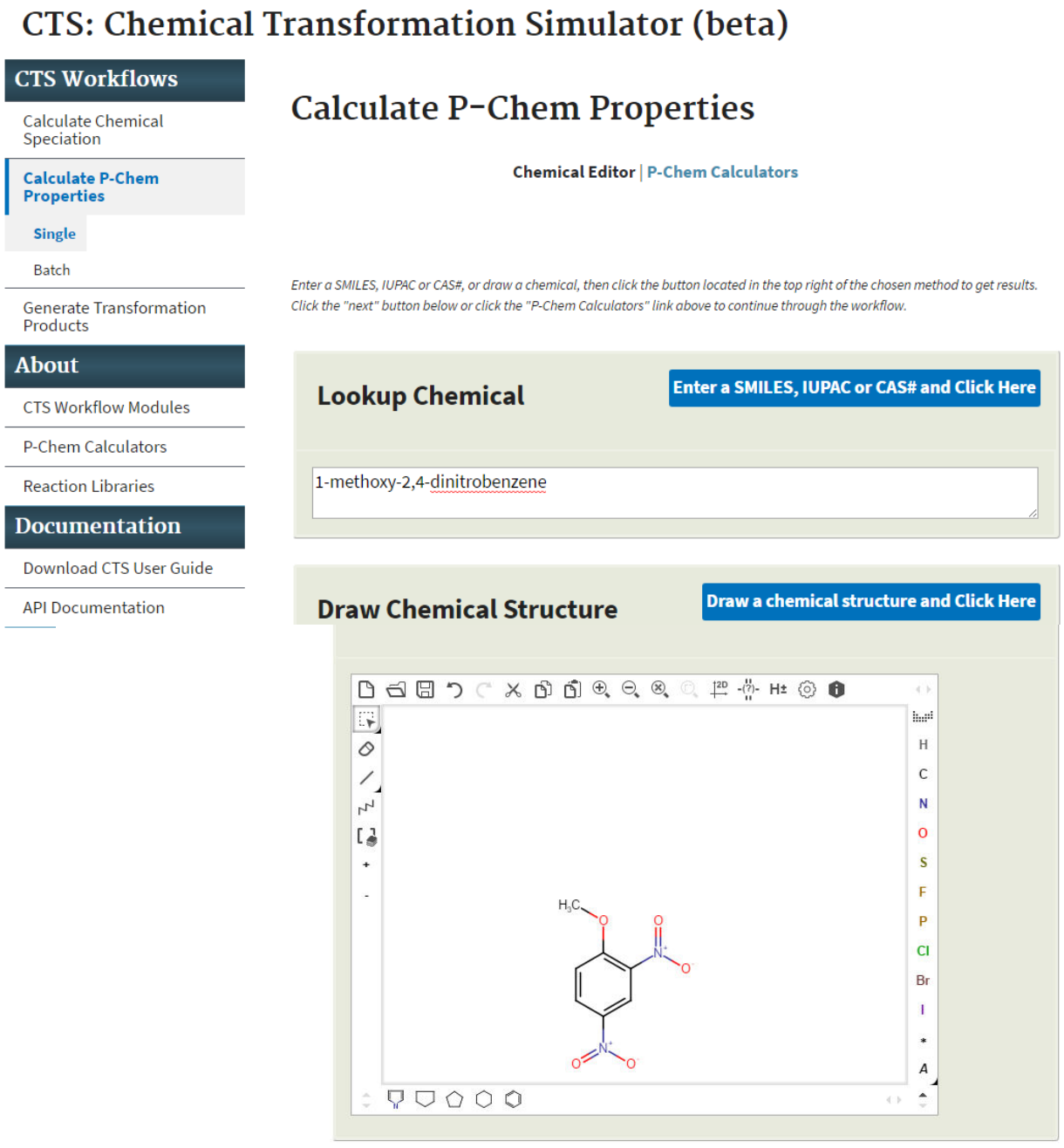
Molecular Information	
smiles	Br[C@@H]1[C@H](Br)[C@H](Br)[C@H](Br)[C@H](Br)[C@@H]1Br
exactMass	551.556978 g/mol
iupac	(1R,2R,3S,4S,5s,6s)-1,2,3,4,5,6-hexabromocyclohexane
mass	557.538 g/mol
formula	C6H6Br6

Calculate p-Chem Properties Workflow

Selection of the Calculate p-Chem Properties Workflow provides this page illustrating the workflow overview. Click on the Go to User Inputs button or the Inputs tab to submit a single chemical for processing, or click on the Batch tab to submit a batch run (currently under construction).



For the purpose of this demonstration, 1-methoxy-2,4-dinitrobenzene has been entered into the chemical editor. Selects the Next button to choose the p-chem calculators and p-chem properties of interest.

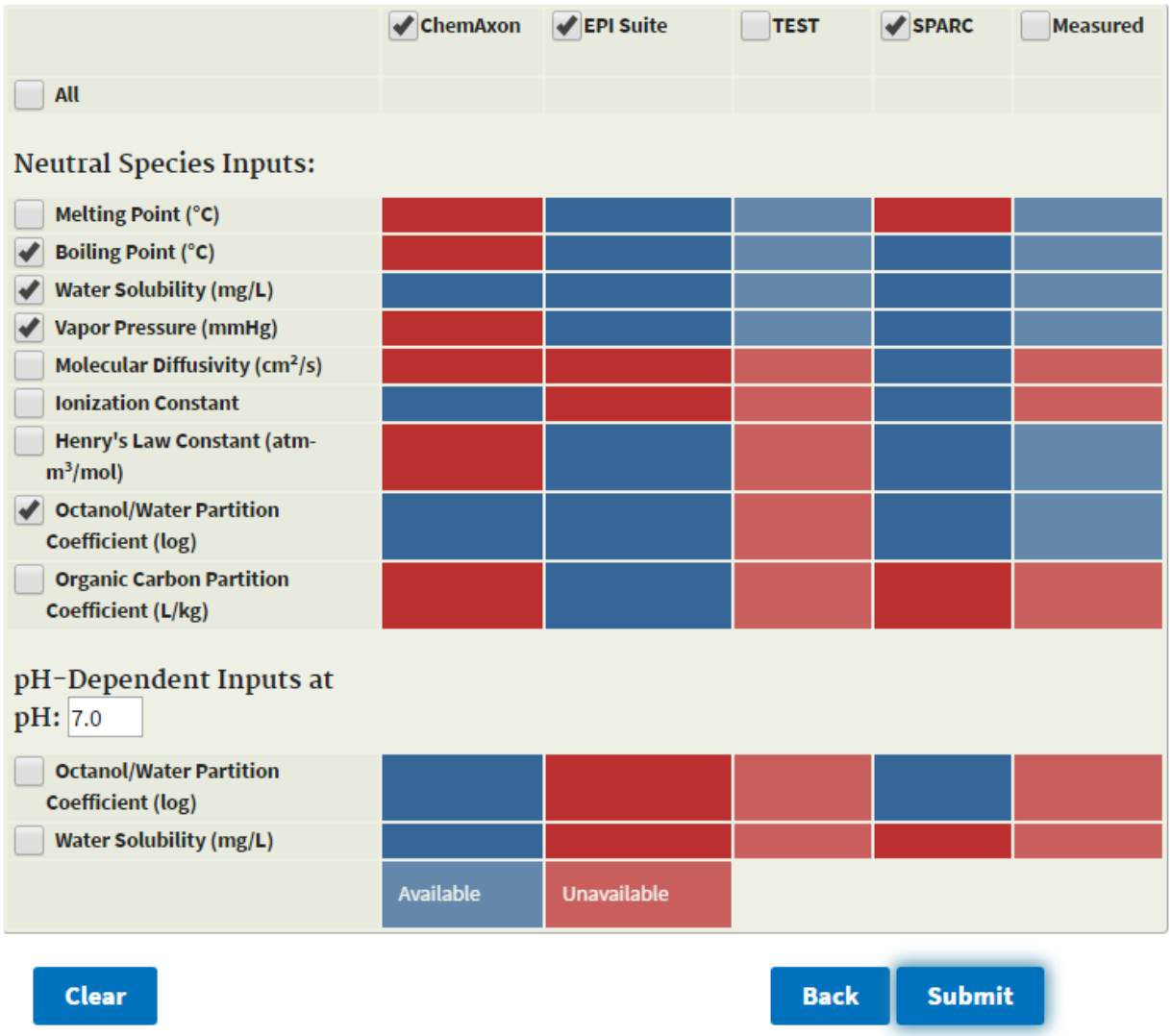


Results

Entered chemical:	1-methoxy-2,4-dinitrobenzene
Initial SMILES:	COC1=C(C=C(C=C1)[N+]([O-])=O)[N+]([O-])=O
SMILES:	COC1=C(C=C(C=C1)[N+]([O-])=O)[N+]([O-])=O
IUPAC:	1-methoxy-2,4-dinitrobenzene
Formula:	C7H6N2O5
CAS #:	
Mass:	198.134
Exact Mass:	198.027671301

Clear Next

Use the Calculate p-Chem Properties Workflow Inputs screen to select p-chem properties and the p-chem calculators of interest. Selection of the All button for the p-chem properties will only provide the available properties for the selected p-chem calculators.



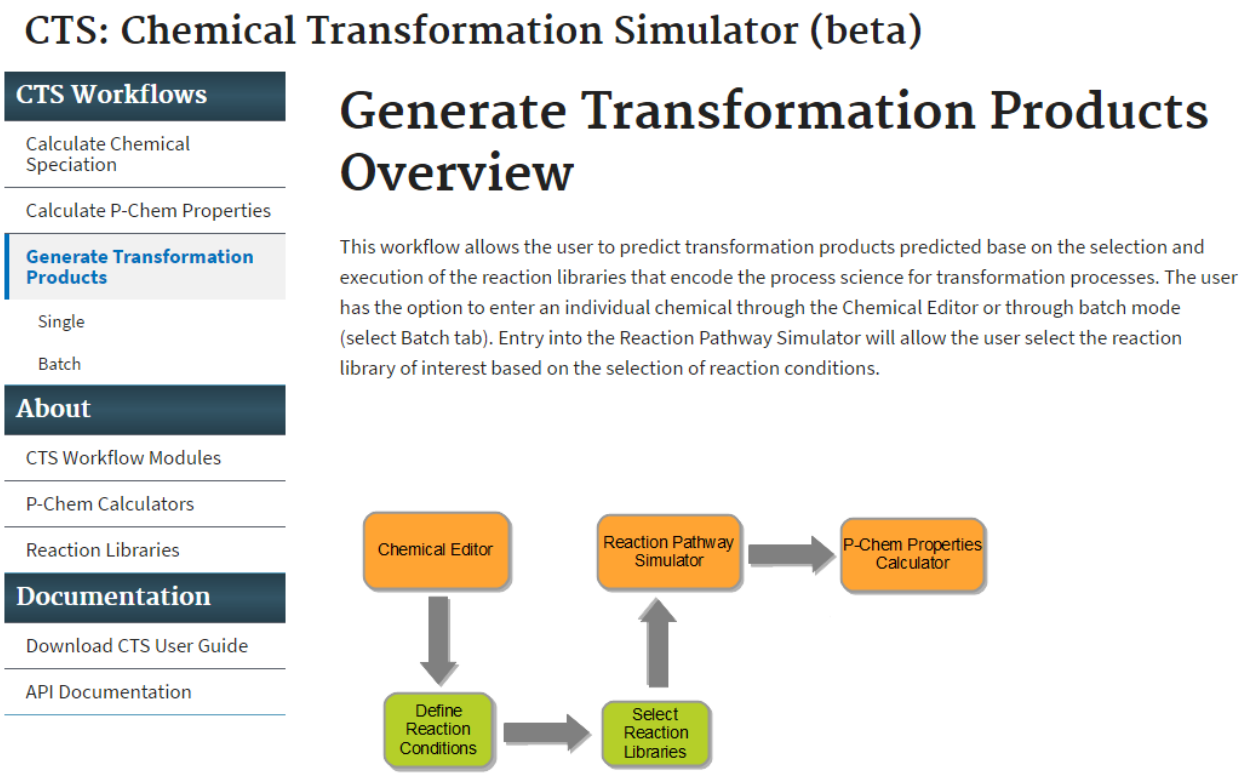
The Calculate p-Chem Properties Workflow Outputs screen provides the results of the previously selected p-chem properties.

- P-Chem Properties Results

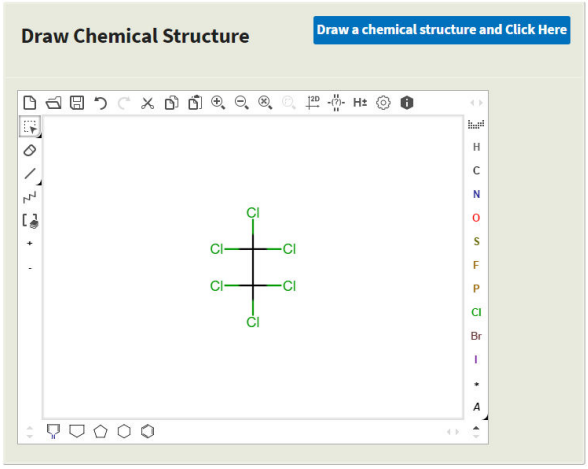
	ChemAxon	EPI Suite	TEST	SPARC	Measured
All					
Neutral Species Inputs:					
Melting Point (°C)					
Boiling Point (°C)		319.62		308.40	
Water Solubility (mg/L)	2.68e+2	6.32e+2		1.27e+3	
Vapor Pressure (mmHg)		1.38e-4		1.36e-3	
Molecular Diffusivity (cm^2/s)					
Ionization Constant					
Henry's Law Constant ($atm\text{-}m^3/mol$)					
Octanol/Water Partition Coefficient (log)	1.74 KLOP 1.70 VG 1.65 PHYS	1.71		2.21	
Organic Carbon Partition Coefficient (L/kg)					
pH-Dependent Inputs at pH: 7.0					
Octanol/Water Partition Coefficient (log)					
Water Solubility (mg/L)					
	Available	Unavailable			

Generate Transformation Products Workflow

Selection of the Generate Transformation Products Workflow provides this window illustrating the workflow overview. Click on the Go to User Inputs button or the Inputs tab to submit a single chemical for processing, or click on the Batch tab to submit a batch run (currently under construction).

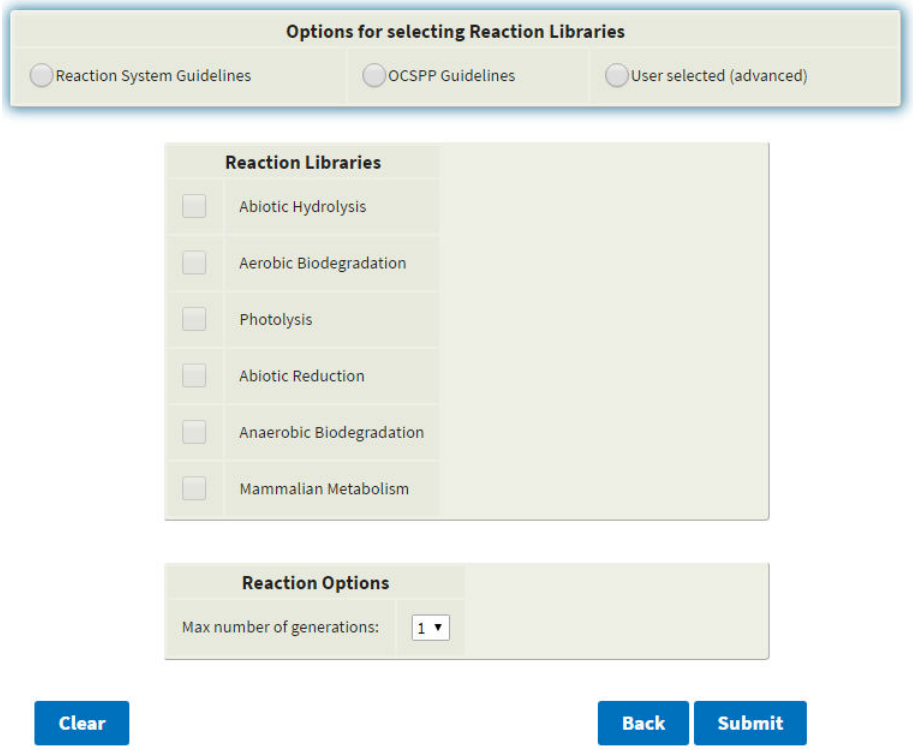


For the purpose of this demonstration, hexachloroethane has been entered into the chemical editor.



The first required input is the selection of the reaction libraries is based on the transformation pathways of interest. Three reaction libraries, including abiotic hydrolysis, abiotic reduction and Phase 1 mammalian metabolism, are available in the β-version of the CTS. A reaction library for photolysis is currently under development, and a seamless linkage to a reaction library for aerobic biodegradation will be available in the next version of the CTS. Three options are available for the selection of one or multiple reaction libraries:

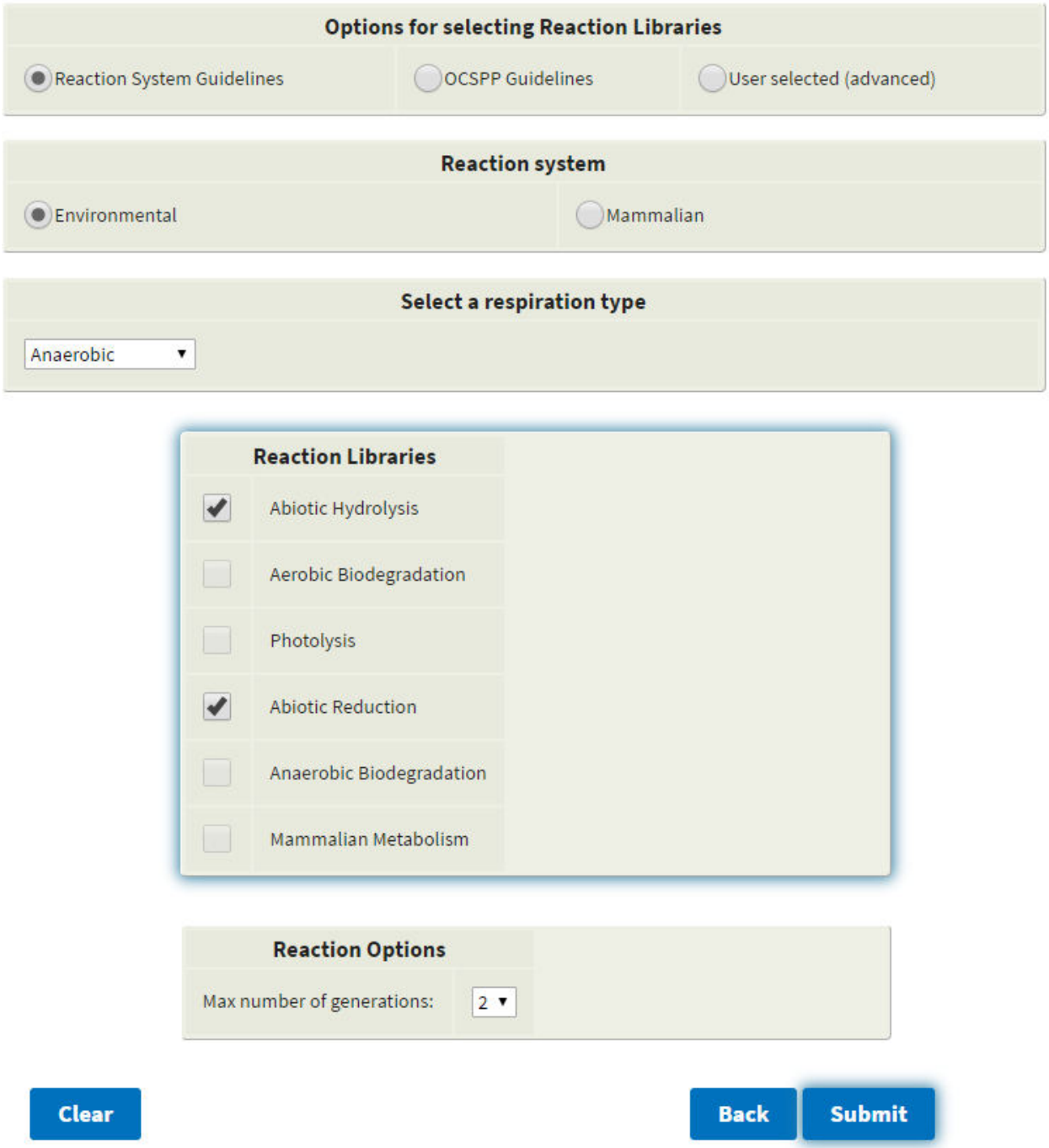
- Reaction System Conditions
- OCSPP Harmonized Test Guidelines
- User Selected (Advanced)



Selection of the Reaction System Conditions provides 2 options for reaction systems: Environmental or Mammalian.

Selection of the Environmental Reaction System provides the option to select respiration type: Aerobic or Anaerobic.

Selection of anaerobic respiration opens the window with the reactions libraries for the transformation pathways that are currently available and will potentially occur under these reaction conditions, which includes abiotic hydrolysis and abiotic reduction.

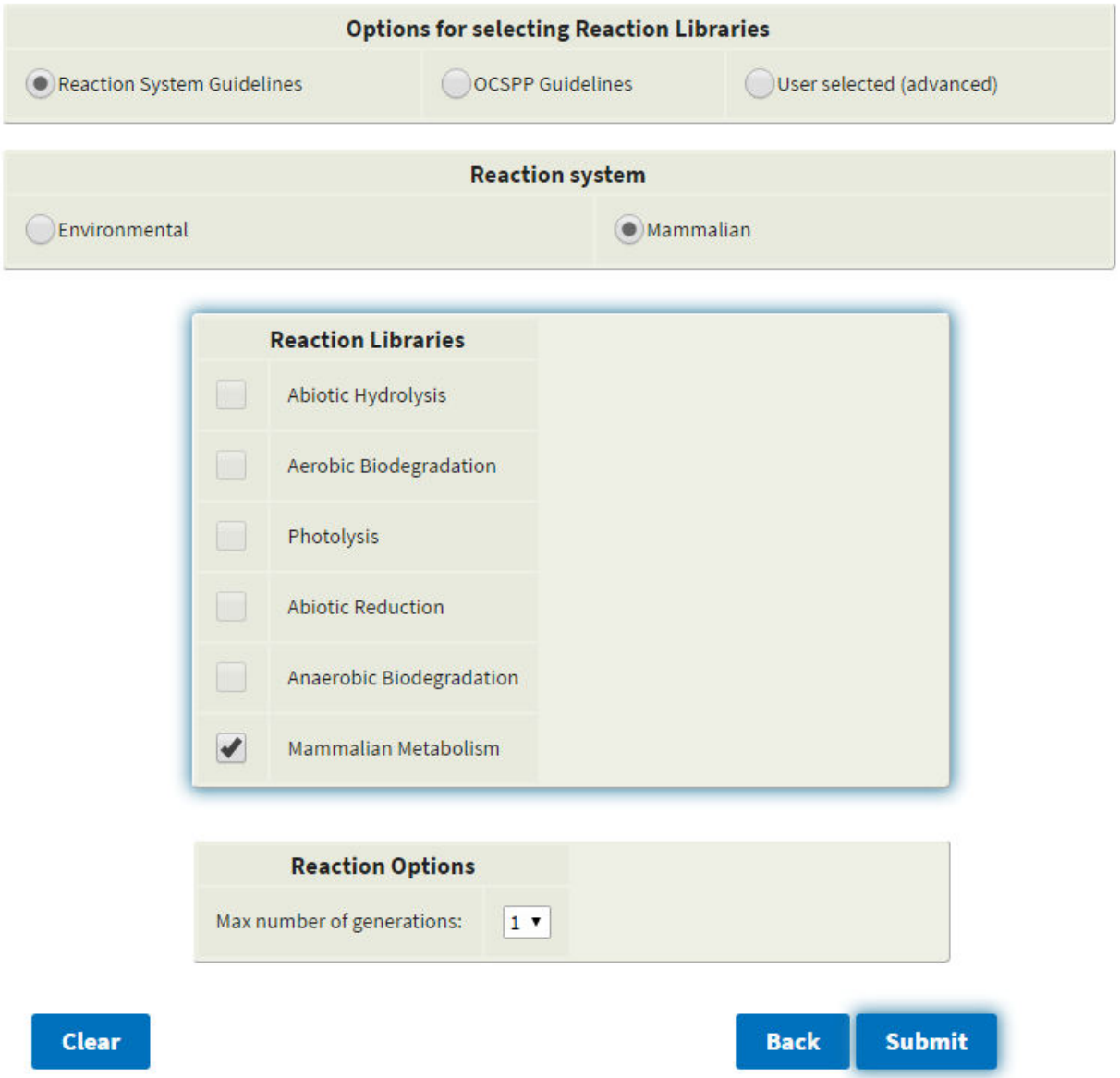


Selection of aerobic respiration opens the window with the reactions libraries that are currently available and will potentially occur under these conditions, which currently includes only abiotic hydrolysis.

Reaction System Guidelines

Selection of Reaction System Guidelines

Selection of mammalian reaction systems opens the window with the mammalian reaction library selected. This is the only option available for the mammalian reaction system.

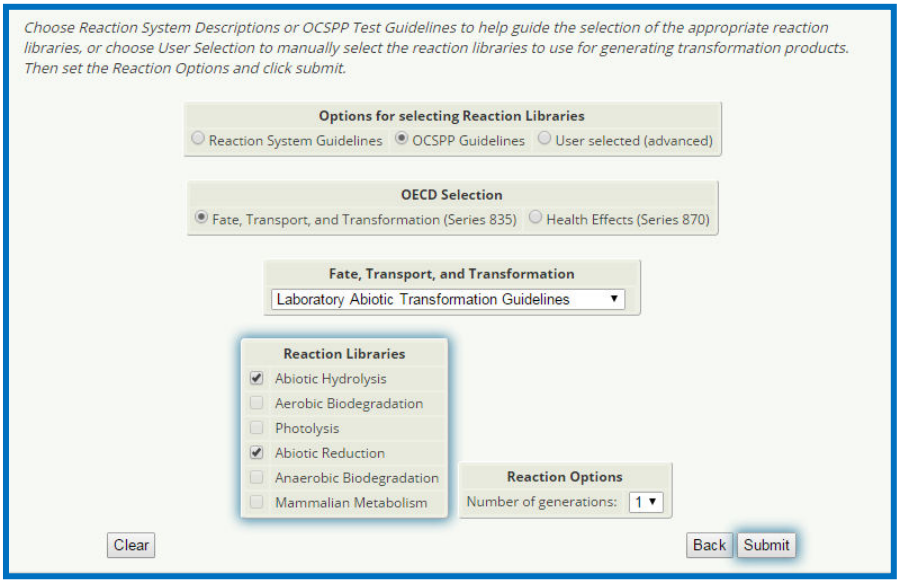


The second option for the selection of reaction libraries is through the selection of the OCSPP s Fate, Transport, and Transformation (Series 835) or Health Effects (Series 870) Guidelines.

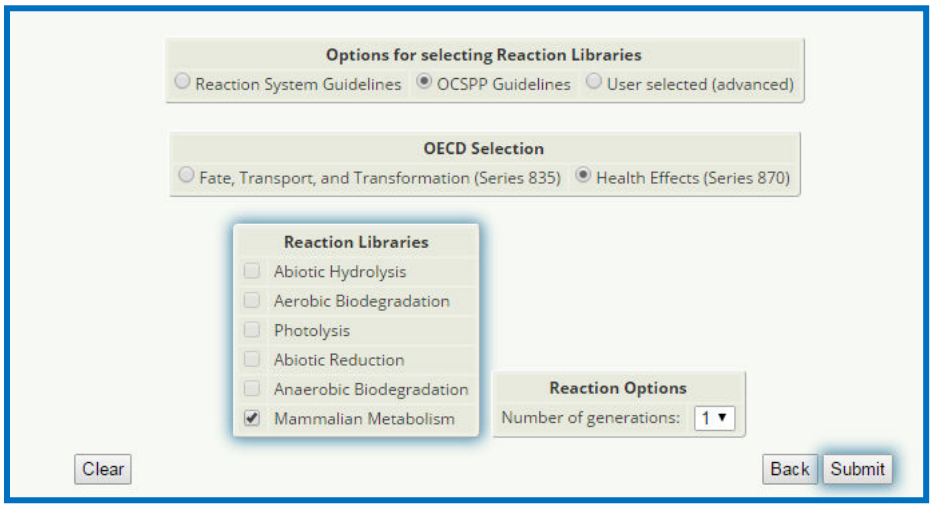
Selection of the Fate, Transformation Series provides three options:

- Laboratory Abiotic Transformation Test Guidelines
- Transformation in Water and Soil Test Guidelines
- Transformation Chemical-Specific Test Guidelines

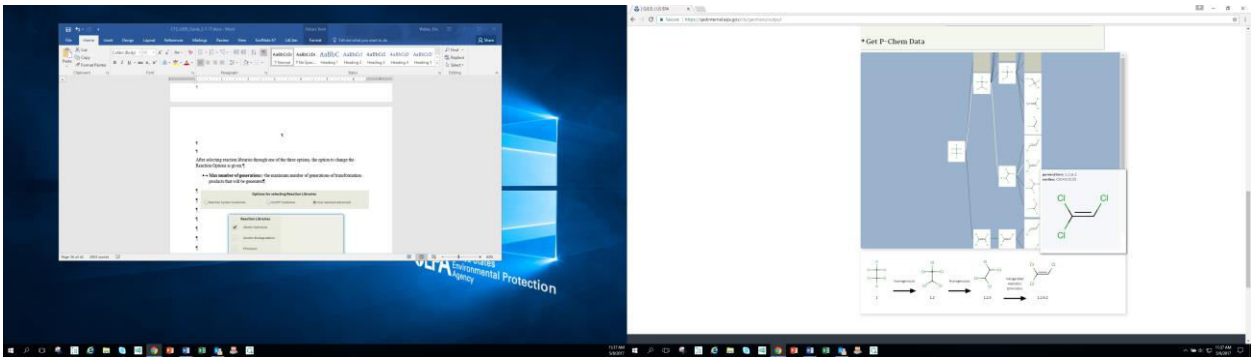
As an example, selection of the Laboratory Abiotic Transformation Test Guidelines shows that both the abiotic hydrolysis and abiotic reduction are appropriate selections for this option.



Selection of Health Effects provides one option for selection of a reaction library (i.e., mammalian metabolism).

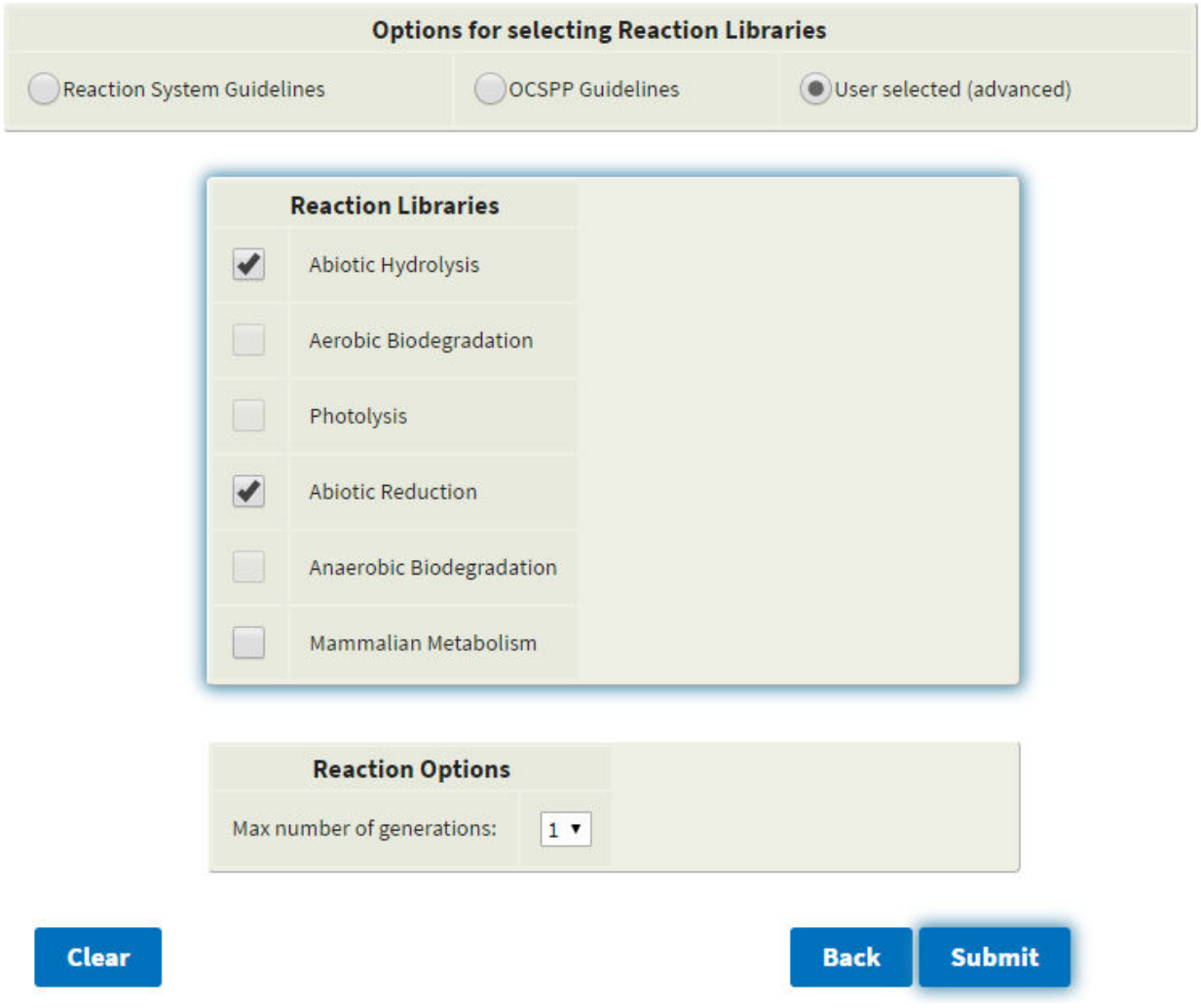


The third option for the selection of reaction libraries is through the selection of the User Selected, which is considered to be an option for the more advanced user. This option provides the ability to select amongst the currently available reaction libraries.

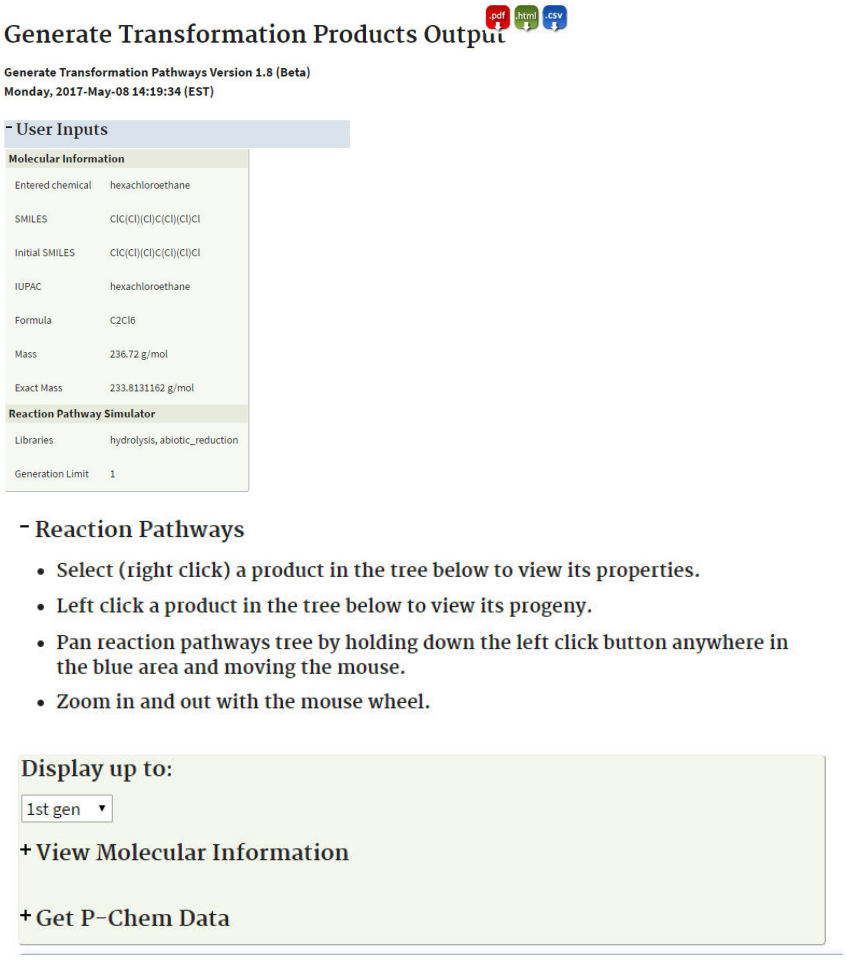


After selecting reaction libraries through one of the three options, the option to change the Reaction Options is given:

- **Max number of generations:** the maximum number of generations of transformation products that will be generated. The default value is set at one.



After selection of the reaction libraries and reaction options have been made, click the submit button to generate transformation products. The results screen summarizes the input data and provides the 1st generation of transformation products (the default value) based on execution of the abiotic hydrolysis and reduction libraries as previously selected.



Generate Transformation Products Output

Generate Transformation Pathways Version 1.8 (Beta)
Monday, 2017-May-08 14:19:34 (EST)

- User Inputs

Molecular Information	
Entered chemical	hexachloroethane
SMILES	ClC(Cl)(Cl)C(Cl)(Cl)Cl
Initial SMILES	ClC(Cl)(Cl)C(Cl)(Cl)Cl
IUPAC	hexachloroethane
Formula	C2Cl6
Mass	236.72 g/mol
Exact Mass	233.8131162 g/mol
Reaction Pathway Simulator	
Libraries	hydrolysis, abiotic_reduction
Generation Limit	1

- Reaction Pathways

- Select (right click) a product in the tree below to view its properties.
- Left click a product in the tree below to view its progeny.
- Pan reaction pathways tree by holding down the left click button anywhere in the blue area and moving the mouse.
- Zoom in and out with the mouse wheel.

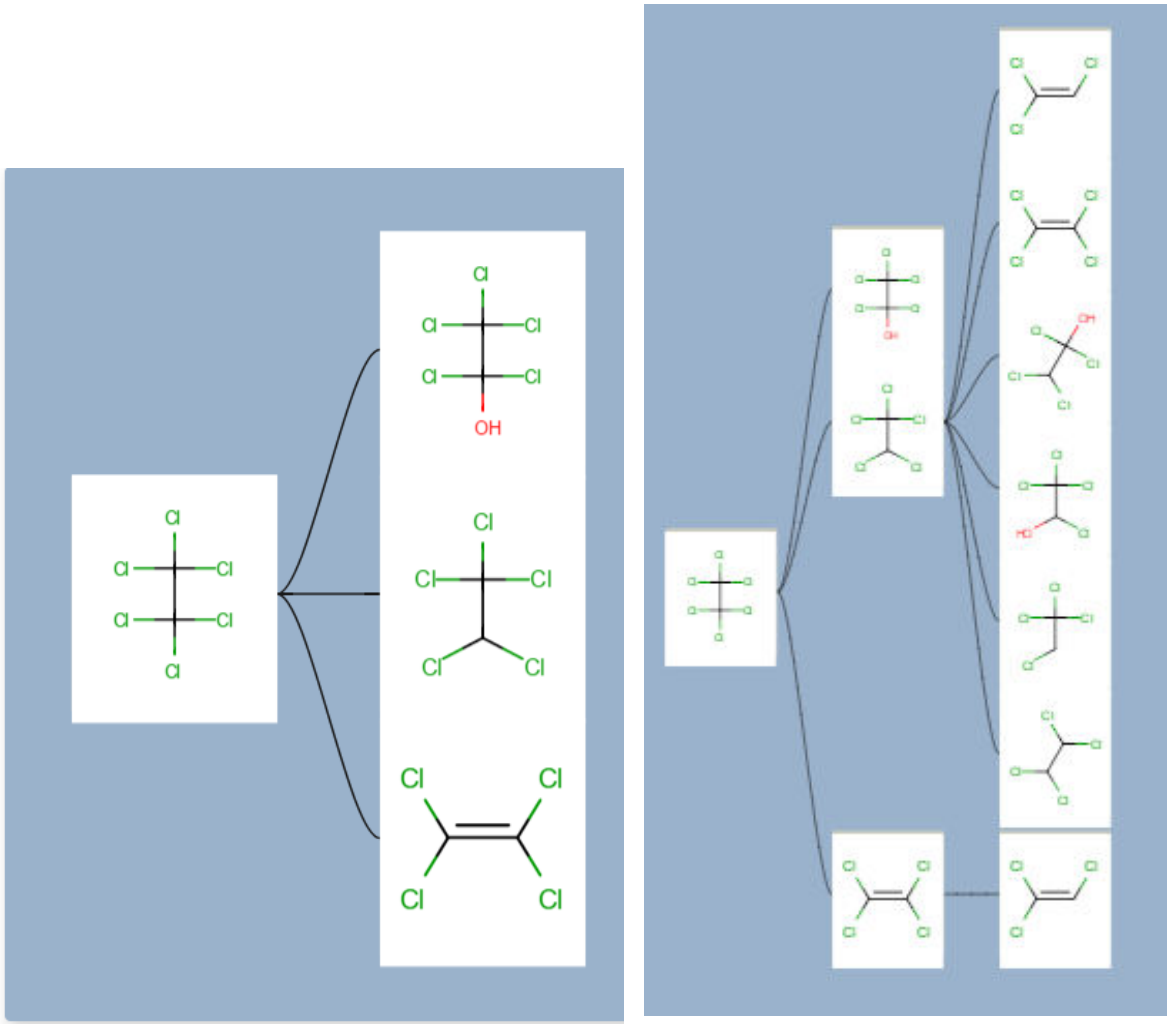
Display up to:

1st gen

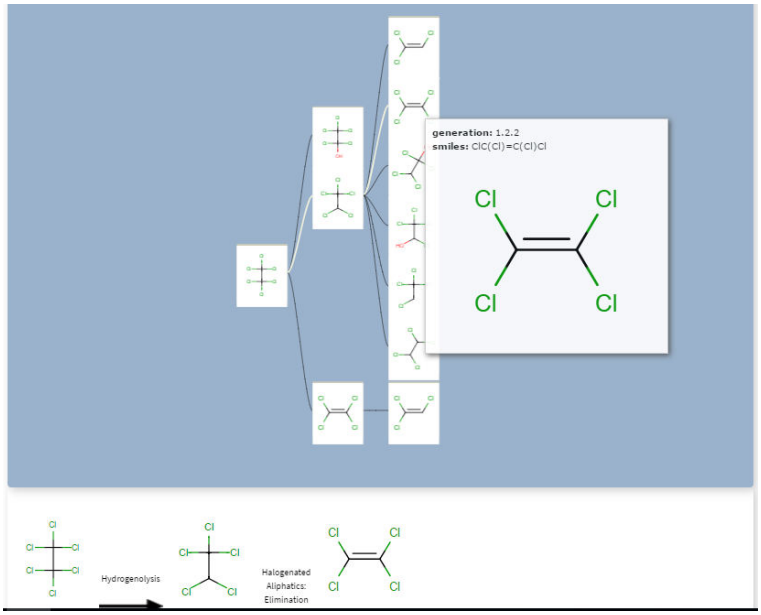
+ View Molecular Information

+ Get P-Chem Data

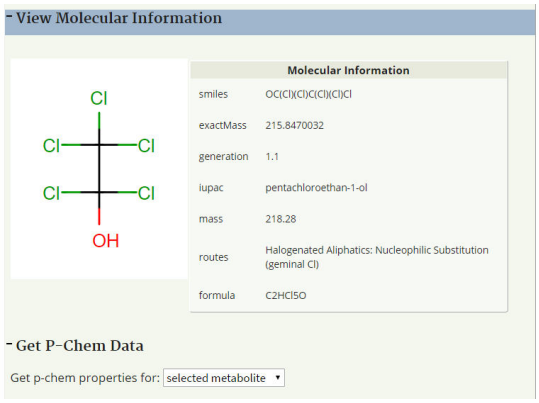
The number of viewed generations can be increased by changing the number of generations in the "Display up to" window. The screen on the left below illustrates the reaction pathway map for the formation of one generation of products. The screen on the right below illustrates the reaction pathway map 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.

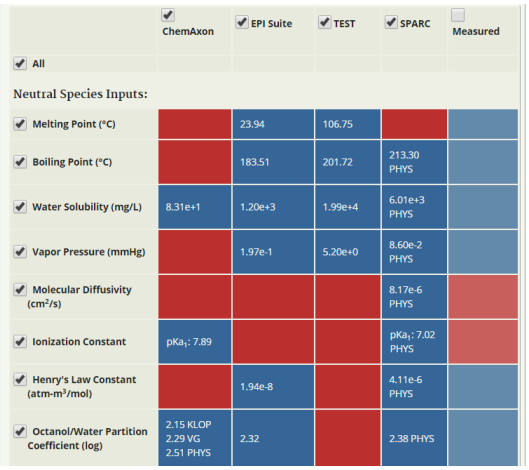


By placing the cursor over a product, a number appears that signifies its place in the reaction pathway map. For this example, tetrachloroethene (1.2.2) is the 2nd product formed in the third generation from the 2nd product (i.e., pentachloroethane, 1.2), which was formed in the second generation from hexachloroethane. By right-clicking on a product, the molecular and metabolite information for the product is displayed. In the β-version of the CTS, the transmissivity, production and accumulation values have no significance. The next generation of transformation products that are predicted to form from a selected product, as well as the reactions that form them, can also be displayed by left-clicking one of the products.



Expanding of the “Get p-Chem Data” menu Reaction Pathways provides the screen on the left below with the various options for p-chem properties and calculators to be applied to the selected transformation product. P-chem properties will be calculated and displayed in the selection table. For example, selection of the All and ChemAxon buttons and then the get data button provides the screen on the right below showing the results for the ChemAxon p-chem calculator for pentachloroethanol.



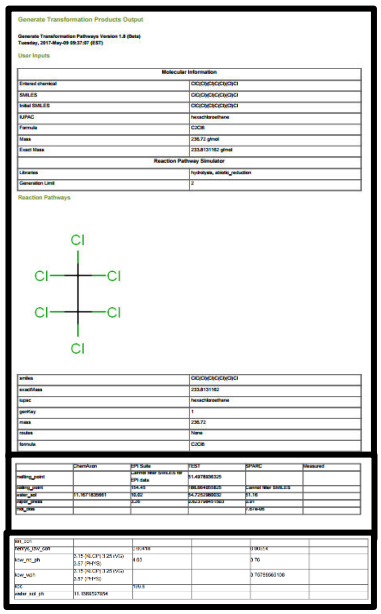


To get p-chem data for multiple metabolites, select the option for up to the first, second, or third generation of metabolites, or for all calculated metabolites from the drop-down menu. Then select the properties and calculators to be used, and click the “Get data” button. The results for multiple metabolites 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 PDF report as described below.

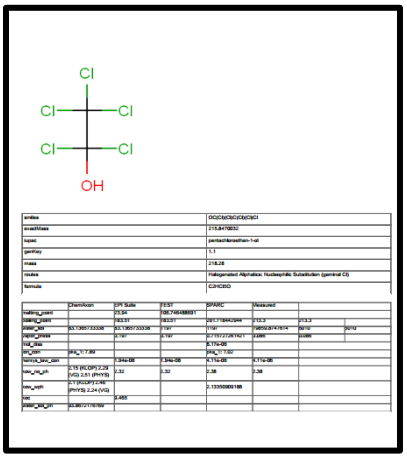
Generation of .pdf , .html and .csv Reports

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 a pdf and html report is shown below.



.PDF report



.HTML report

The .CSV report is generated in a tabular format as shown below:

	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)(	1,1,2-trich	C2HCl3	131.38	129.9144

.CSV report