



Biotransformation Susceptibility (BiotS) software

This software was developed by the Central Science Laboratory under The Long-Range Initiative funding from the European Chemical Industry Council (CEFIC) during project Eco6.

The software allows the user to enter the structure of a query molecule into the software and then search for the presence of susceptible fragments. If susceptible fragments are identified the user can access the location of these fragments to make a decision on their likelihood to biotransformation, identify within which aquatic taxa the biotransformation has been observed and access biotransformation study information, AMBIT descriptors, metabolic pathway and bioproperty data for chemicals where the susceptible fragments have been observed metabolising within publicly available literature. A detailed description of the software is provided below:

BiotS development and structure

BiotS was developed as a standalone desktop application with a graphical user interface (GUI) in the programming language Java version 5. Java is platform independent so this enables the software to be installed on Windows, Unix, Linux, AppleMacintosh™ and Solaris operating systems. The final fragment library and additional parameter data are embedded into the user interface of BiotS. However BiotS has been designed in such a way that it is possible to easily update the embedded data so the analysis can incorporate new data when available. These new data could be either additional aquatic biotransformation data or data relating to the biotransformation/metabolism of chemicals in new systems e.g. mammals or microbes. An in-built help system has also been incorporated into BiotS to allow users to learn how to use the software.



Installation

The software should install directly from the installation file supplied (BiotS_1_4.exe). However, there is a requirement that the installation computer should be running a version of Java that is version 5 or higher. BiotS will not install until an appropriate version of Java is running on the computer. A recent version of Java can be freely downloaded from the Java website:

http://www.java.com/en/download/windows_manual.jsp

Interface information windows

The general software user interface has five main windows that provide query molecule search results (1-3) and supporting data on studies where fragment biotransformation has been observed (4 & 5) (Figure 1): (*N.B. In the latest version of the BiotS, windows 2 and 3 have been switched to better display the information*)

The screenshot displays the BiotS software interface with five numbered windows:

- Window 1:** A table listing SMILES strings and their corresponding Alurns and No. of Res. Biot. values.
- Window 2:** A chemical structure diagram of a chlorinated aromatic compound with a phosphate group.
- Window 3:** A table with columns: Taxa, Reactio., Group, No. of S., and HIT ID.
- Window 4:** A table with columns: TestID, Species, Common Name, Radiolabel study, Water type, and other parameters.
- Window 5:** A table with columns: Compound, TestID, Type, Operator, Value, Upper, Lower, Std dev, units, Comment, and BioassayID (ChemID).

Figure 1. Five windows used to display biotransformation data within BiotS



1. This *Fragment window* reports on any fragment hits found within a query molecule and provides the user with the fragment ID number, SMILES notation of the fragment, number of atoms in the fragment and the number of metabolic steps that fragment has been observed to undergo biotransformation.
2. This *Structure window* provides the user with a two-dimensional representation of the query molecule and when a fragment is highlighted in window 1 (by clicking the left button of the mouse on the fragments row) the position of the fragment is displayed within the query molecule structure.
3. When a fragment is highlighted then the *Taxa window* (window 3) displays the taxa within which the chosen fragment has been seen to biotransform. This window provides the user with taxa (e.g. ray-finned fish, decapoda), reactions type (e.g. addition, loss, substitution) and group involved (e.g. glucuronic acid).
4. The *Study window* (window 4) provides the user with access to experimental data concerning the chemical within which the chosen fragment has been seen to biotransform. If one of the taxa are selected in window 3 then the user can access information on species, environmental media, parent/exposure chemical, exposure and elimination duration and the original reference. When a line within this window is selected with a right mouse click the user can further access data relating to the biotransformation pathway, AMBIT molecular descriptors and the full bibliography reference for the study (this is explained in more detail below).
5. The *Bioproperty window* (window 5) provides the user with access to bioproperty data relating to the study and chemical selected in the *Study window*.

Once a query is performed, depending on the users requirements the size of the windows can be adjusted to allow better examination of the results. Small arrows



situated between the windows e.g. between *Fragment window* and *Structure window* highlighted in circle in Figure 1, can be used to adjust the display. The arrows can be used to display just the *Fragment window* (Figure 2), or will hide the *Fragment window* and display the *Structure window* and an enlarged *Taxa window* (Figure 3).

HllC	SMILES	Alutms	No. of Reaction Steps
9	CO	2	26
20	COO	3	17
283	CC	2	26
488	c1ccccc1	6	3
749	C=CC	3	5
750	C=C	2	4
1284	P(O)OC	5	12
1285	P(O)(O)OC	6	12
1284	POC	3	12
1285	P(O)OC	4	12
1286	P(O)(O)OC	5	12
1322	P(O)(O)C=O	6	3
1323	P(O)(O)(O)C=O	7	3
1330	P(O)C=O	4	3
1331	P(O)(O)C=O	5	3
1332	P(O)(O)(O)C=O	6	3
1632	c1ccc(cc1)CO	8	2

Figure 2. Window display adjusted to view the Fragment window

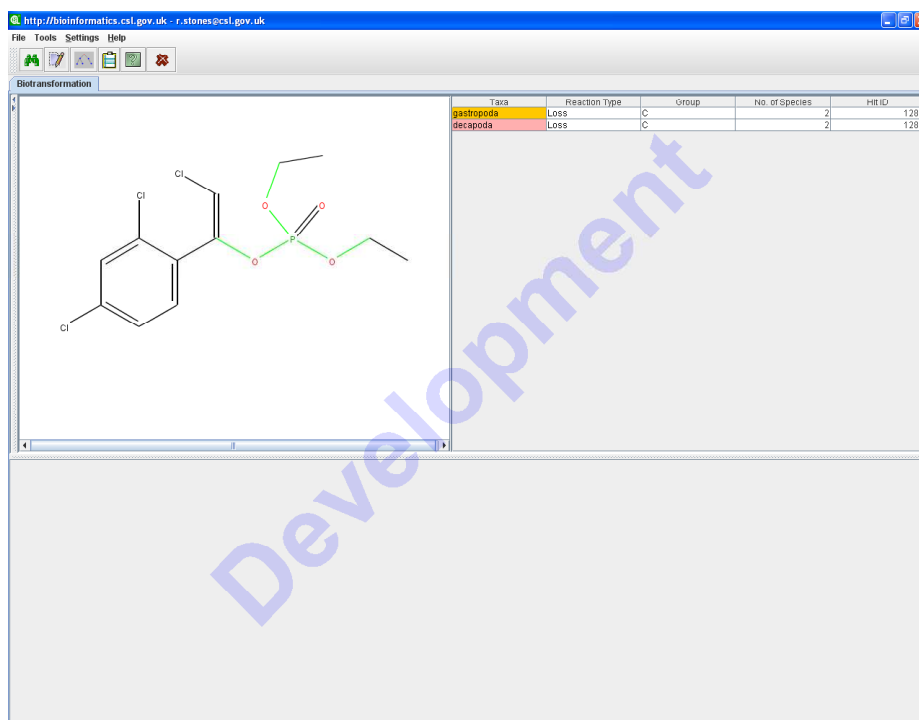





Figure 3. Window display adjusted to view Structure and Taxa window


Functionality


Within the software the functionality can be accessed using either menus, the toolbar or shortcut keys. Detailed below is a simple description on how to select each function. A more complete description of the functionality can be found below (from Section 0 onwards).


- Open the query molecule input window
 - On the **File** menu, click **EnterQuery**
 - Press the binocular button 
 - Press **Ctrl+Q**
- Exit the software
 - On the **File** menu, click **Quit**
 - Press the cross button 
 - Press **Alt+Q**




- View bibliographic reference for supporting information
 - On the **Tools** menu, click **Reference**
 - Press the pen and paper button 
 - Press **Ctrl+R**

- View biotransformation pathway for chemicals that contain fragments
 - On the **Tools** menu, click **Pathway**
 - Press the pathway button 
 - Press **Ctrl+P**
 -

- View AMBIT molecular descriptors for study chemicals
 - On the **Tools** menu, click **MolDescriptors**
 - Press the pathway button 
 - Press **Ctrl+M**

- Generate a summary report for fragments in query molecule
 - On the **Tools** menu, click **Summary**
 - Press the clipboard button 
 - Press **Ctrl+S**

- Access the user help section
 - On the **Help** menu, click **Help**
 - Press the question mark button 
 - Press **Ctrl+S**

- Optimise the software for Microsoft Windows operating system
 - On the **Setting** menu, click **Look & Feel** and select **Default**
(software restart needed for change to take effect)



- Optimise the software for Macintosh operating system (Mac OS X)
 - On the **Setting** menu, click **Look & Feel** and select **KDE**
(software restart needed for change to take effect)
- Set the default directory for files (.smi. structural and .txt reports)
 - On the **Setting** menu, click **SetOpenDir**
 - Press **Alt+D**

Results from a query molecule

Query Molecule Entry

Once the BiotS software has been started the user can submit a query molecule using the query molecule entry window (Figure 4). There are three methods for entering query molecules

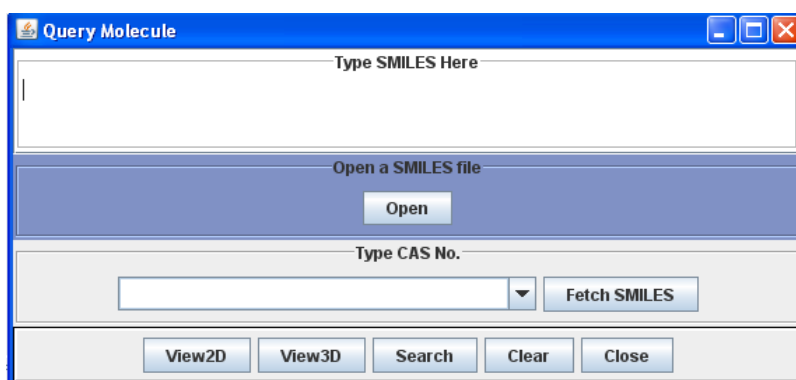


Figure 4. The query molecule entry window

1. SMILES notation can be typed directly into the window or pasted in using **Ctrl+V**.
2. A SMILES file (.smi) can be used to import the SMILES notation ('Open a SMILES file') by pressing the **Open** button and selecting the desired file.
3. A CAS number can be selected from the compounds already in the dataset and used to access the SMILES notation for that chemical ('Type a CAS no.')



Running a search

Once a SMILES notation has been entered the user can either view the molecule in 2D or 3D (which allows structure rotation) by pressing the **View 2D** and **View 3D** buttons respectively or search the query molecule for susceptible fragments by pressing the **Search** button. The software performs two consecutive searches once a query molecule has been entered:

1. Search for an exact match for the query molecule in the dataset i.e. is the metabolism/biotransformation of the query molecule already present within the collated dataset.
2. A search is undertaken to identify whether any of the 'susceptible fragments' in the fragment library are present in the query molecule (the time to complete this search depends on the size of the query molecule);

Search methodology

Once the query is started then the entered structure is searched using a maximal common substructure determination algorithm, this runs against the database to see if there is an exact match to a chemical in the database. If there is no exact match the software searches the fragment library using a molecular fingerprint match (Figure 5) (this is not substructure matching). The initial molecular fingerprint matching is far more rapid than using substructure matching. The molecular fingerprint matching can make incorrect matches but it reduces the search space dramatically because unlike the chemical database search we are searching thousands of molecular fragments. The hits (molecular fragments) from the molecular fingerprint search are then subjected to the more time consuming maximal common substructure, where only complete 100% matches with fragments in the library are reported. Once molecular fragments have been matched using the maximal common substructure they are displayed within BiotS along with supporting information.

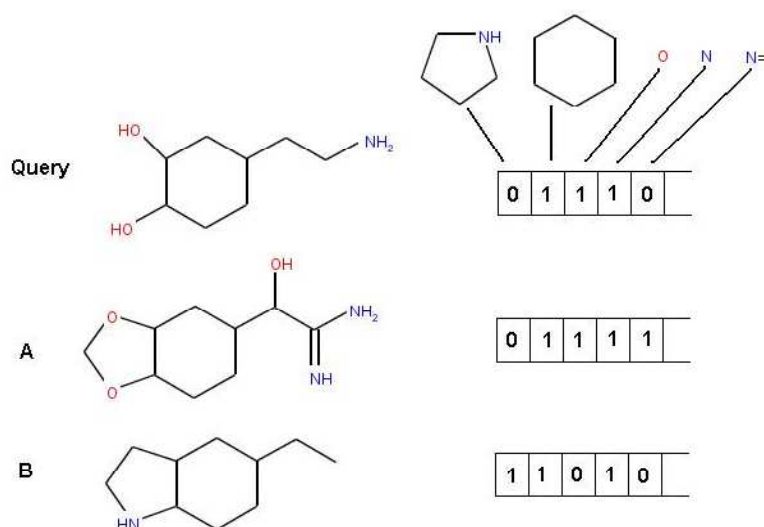


Figure 5. Bitstring representation of a query substructure with bitstrings of two database molecules. Molecule A passes since all bits set to 1 in the query are also set in its bistring. Molecule B, fails since bit representing presence of O (oxygen group) is set to 0 in its bistring.

Susceptible fragment results in query molecule

Once the searching has finished the results are initially displayed within the *Fragments window*, *Structure window* and *Taxa window*. As an example to demonstrate the output, the organophosphorus insecticide chlorfenvinphos has been used as a query molecule within the software and the results are displayed in Figure 6.

In the *Fragments window* it can be seen that seventeen ‘susceptible fragments’ have been identified within the molecule (Figure 6). At present BiotS does not provide the user with any measure of the biotransformation susceptibility of a fragment identified within a query molecule. The results should be interpreted as a fragment has been identified in the query molecule that has previously been observed to undergo biotransformation within one or more aquatic taxa. Moreover the output requires some user interpretation as some fragments may be completely buried within the query molecule or at least have their reaction centre buried within the query molecule and therefore be unavailable for the identified biotransformation. When examining the results it is also obvious that fragments can be grouped together i.e. they have the



same reaction centre but progressively increase in size by one or more atoms. The most appropriate fragment from this perceived group can then be selected by the user either on the basis on the largest fragment by number of atoms (in the 'Atoms' column of *Fragment window*) or by the number of metabolic steps the fragment has been observed to biotransform within (in the 'No. of metabolic steps' column in the *Fragment window*).

If the *HitID* column for a specific fragment is highlighted in green then this indicates that this fragment is in fact an exact match for the query compound i.e. the biotransformation data that can be accessed in the subsequent windows (see below) relates directly to the biotransformation of the query molecule.

The screenshot shows a web application interface for biotransformation search. The main window is titled "Biotransformation" and contains a table of results. The table has columns for HitID, SMILES, Atoms, and No. of Reaction. The results are as follows:

HitID	SMILES	Atoms	No. of Reaction
9	CO	2	20
20	CCO	3	17
293	CC	2	26
488	c1ccccc1	6	3
749	C=CC	3	5
760	C=C	2	4
1264	P(O)(OC)OC	5	12
1265	P(O)(OC)OC	6	12
1284	POC	3	12
1285	P(O)OC	4	12
1286	P(O)(O)OC	5	12
1322	P(O)(OC)=O	5	3
1323	P(O)(OC)(OC)=O	7	3
1330	P(O)=O	4	3
1331	P(O)(O)=O	5	3
1332	P(O)(O)(O)=O	6	3
1632	c1cc(ccc1)OO	8	2

The chemical structure viewer shows the structure of chlorfenvinphos, which is a phosphorus-containing insecticide. The structure is a 2,4-dichlorophenyl ring attached to a vinyl group, which is further attached to a phosphorus atom. The phosphorus atom is bonded to two ethoxy groups and one oxygen atom. A red circle highlights the "gastro" reaction type in the summary table.

Reaction	Group	No. of S.	HitID
gastro	Loss	2	1323

Figure 6. An example of the results of query molecule searching for chlorfenvinphos



When a specific fragment is selected (by clicking on the row within the *Fragment window*), the position of the fragment within the query molecule will be displayed in the *Structure window* and the taxa the fragment has been seen to biotransform within is displayed in the *Taxa window* together with the reaction type and the moiety involved in the reaction. The extent of data/taxa displayed obviously depends on the number of organisms this fragment has been seen to biotransform in. For example in Figure 6 it can be seen that the selected fragment (P(OC)(OC)=O) demonstrates demethylation (loss of C) within a gastropod. The taxa currently present in the dataset include ray-finned fish, echinoderms, aquatic macrophytes, gastropods, decapods and crustaceans.

The screenshot shows the bioinformatics software interface. The top window is titled "Biotransformation" and contains a table with columns: HitID, SMILES, Atoms, and No. of Reaction. The table lists various chemical fragments and their associated reaction counts. A chemical structure of a complex organic molecule is displayed in the center, with a red circle highlighting a specific fragment. Below the structure is a table titled "Loss of C in gastropods" with columns: TestID, Species, Common Name, Radiolabel study, Water type, Exposure chemical, Exposure route, Exposure concentr., Exposure duration, Elimination duration, and Reference. The table lists two entries: 129/Dipangopaludina (Japanese mystery) and 130/Physa acuta (Sewage or Bladder). Below this is a table titled "Compound" with columns: Compound, TestID, Type, Operator, Value, Upper, Lower, Std dev, units, Comment, BioprodID, and ChemID. The table lists various compounds and their associated test results.

Figure 7. Available supporting information in the Data window

If the gastropod is selected (by clicking on the row within the *Taxa window*) then supporting information is provided in the *Study window* below on the experimental



study that investigated the biotransformation (see Figure 7). In the *Study window* information relating to a specific study where the chosen fragment has been seen to biotransform in an aquatic organism can be viewed. In the example the fragment identified in chlorfenvinphos has been seen to biotransform within the organophosphorus insecticide fenitrothion. The supporting information highlighted indicates that a radio-labelled fenitrothion biotransformation study has been performed in the freshwater snail *Cipangopaludina japonica* and exposure was at 0.1 mg/L for 3 hours with a depuration phase of 6 hours. The final column of this section also provides the user with the literature reference for the study (a full bibliographic reference can also be accessed see section). Moreover if the specific study is selected in the *Study window* (by clicking on the row within the *Study window*) then bioproperty data relating to the study compound and the study are presented in the *Bioproperty window* (Figure 7). Where a compound is highlighted in green it indicates that the bioproperty data relates specifically to the exposure chemical from the study whilst non-highlighted chemicals provide bioproperty data for metabolites of the exposure chemical (as determined in the chosen study). In addition to the specific study bioproperty data, BCF data from the BCF Fish Gold Standard Database (CEFIC project Eco7) are also displayed for the study chemical if available.


For the example study in Figure 7, four BCF Fish Gold Standard Database values are provided together with fenitrothion data relating to biological half-life, elimination half-life and BCF and TCF specific to the snail *Cipangopaludina japonica* as well as metabolite formation data in the organism and depuration water.

Biotransformation study data

Additional information can be accessed for a biotransformation study of a chemical displayed in the *Study window* by selecting the row of interest in the *Study window* with a right mouse click. This will display a list with three options; Biotransformation pathway, Reference and Molecular descriptors. These options can also be selected by highlighting the row in the *Study window* by clicking the left-hand mouse button and then pressing the respective button on the Toolbar (see section 0).



Biotransformation pathway

By selecting *Biotransformation pathway* (or the button ) , if available, this will provide the user with the biotransformation pathway (within a new window) proposed within the experimental study of interest. For example Figure 8 shows the biotransformation pathway of fenitrothion within the gastropod *Cipangopaludina japonica*.

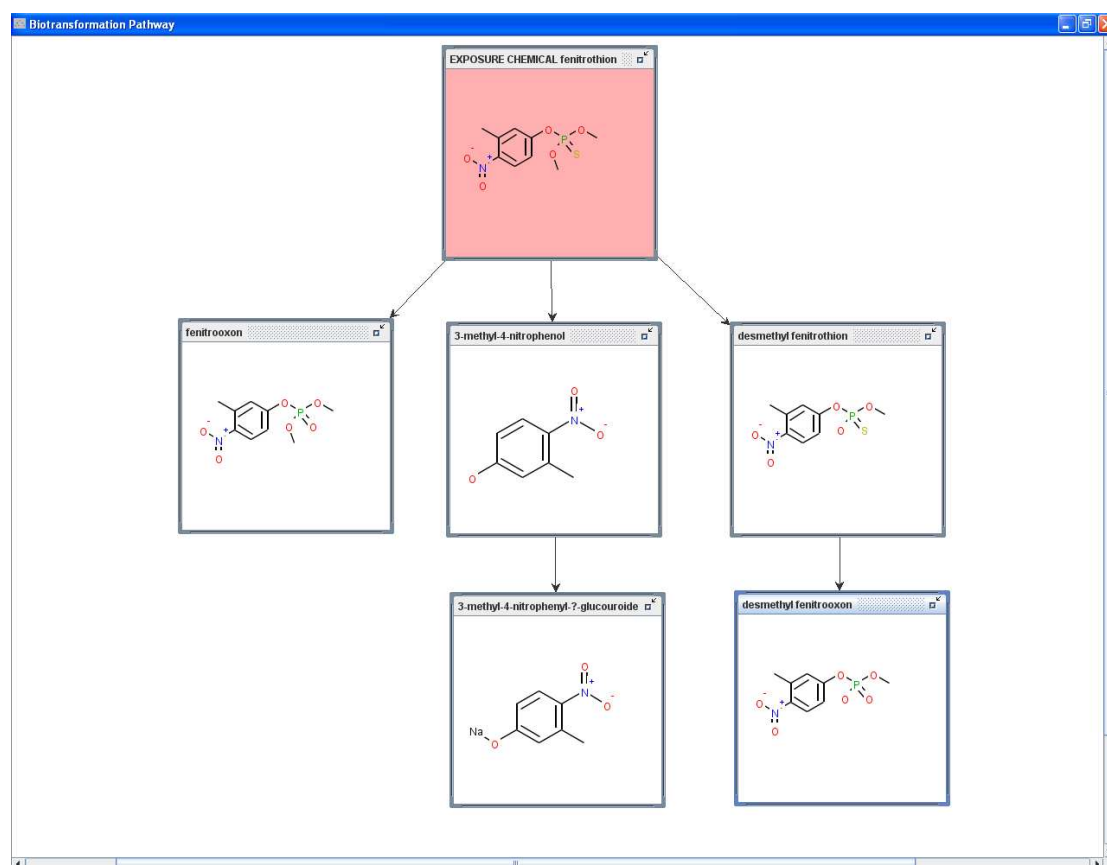

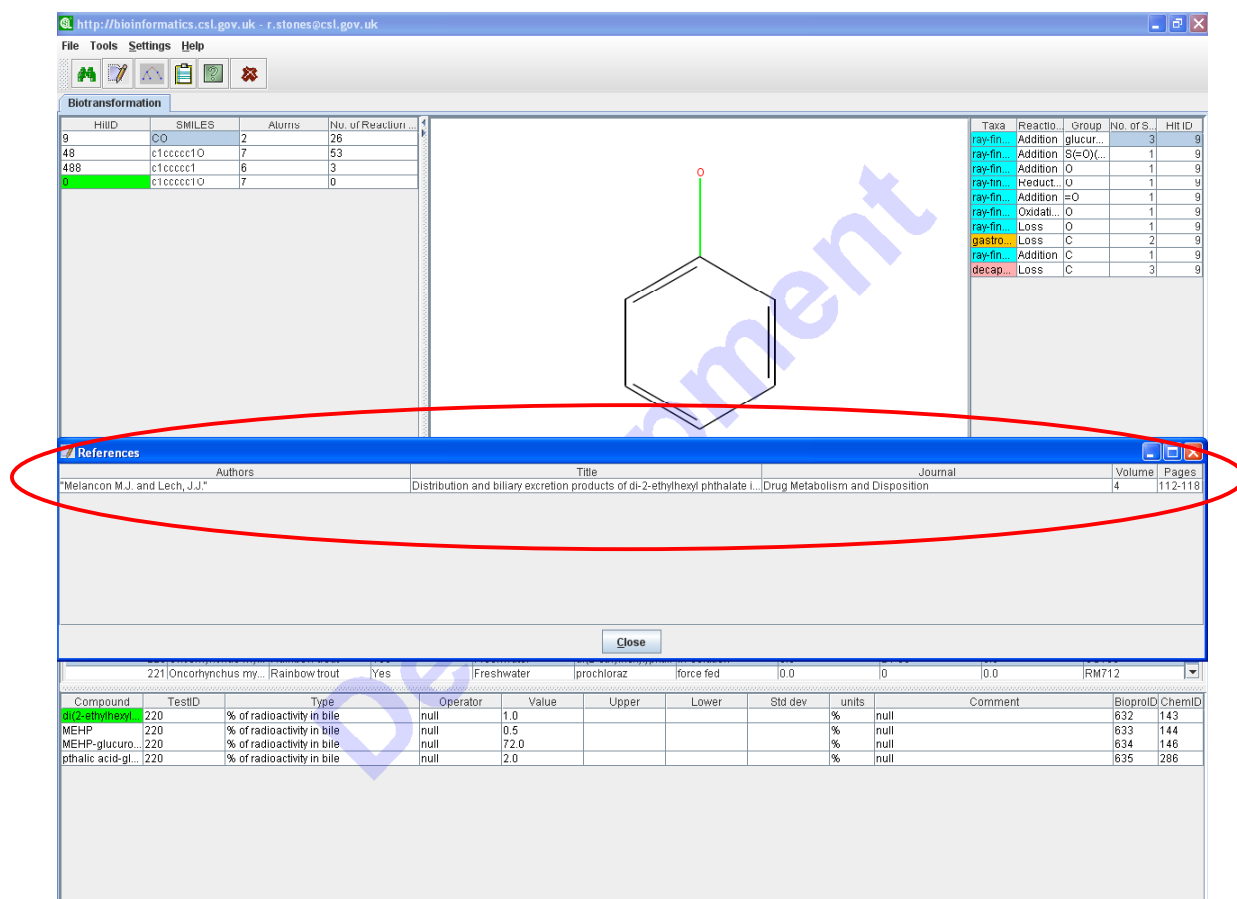


Figure 8. The biotransformation pathway of fenitrothion



Bibliographic reference

By selecting *Reference* (or the button ) this will provide the user with the full bibliographic reference (within a new window) of the experimental study of interest (Figure 9).



The screenshot shows a software interface with a 'Biotransformation' window and a 'References' window. The 'References' window is highlighted with a red oval and contains the following bibliographic reference:


Authors	Title	Journal	Volume	Pages
Melancon M.J. and Lech, J.J.	Distribution and biliary excretion products of di-2-ethylhexyl phthalate I.	Drug Metabolism and Disposition	4	112-118

Below the 'References' window, there is a table with the following data:

Compound	TestID	Type	Operator	Value	Upper	Lower	Std dev	units	Comment	BioassayID	ChemID
di-2-ethylhexyl	220	% of radioactivity in bile	null	1.0				%	null	632	143
MEHP	220	% of radioactivity in bile	null	0.5				%	null	633	144
MEHP-glucuro...	220	% of radioactivity in bile	null	72.0				%	null	634	146
phthalic acid-gl...	220	% of radioactivity in bile	null	2.0				%	null	635	286

Figure 9. The full bibliographic reference for a biotransformation study

Molecular descriptors

By selecting Molecular descriptors (or the table button ) this will provide the user with molecular descriptors from the AMBIT database (Jeliazkova, 2007), within a new window, for the compound within the experimental study of interest (Figure 10). (The descriptors currently within the software where provided by the CEFIC RLT for the compounds within the database).



CAS No.	Common_name	MW	XLogP	TPSA	CrossSectionalDiameter	MaximumDiameter	NSC	RotatableBondsCount
389-08-2	nalidixic acid	232.0848389	2.269	70.5	7.4536	9.8891	82174	5

Figure 10. The AMBIT molecular descriptors display for a study compound

Summary report

Once the user has investigated the identified susceptible fragments and all the available supporting experimental biotransformation information for a query molecule they can generate a summary report. The report provides the user with the 2D structure of the query molecule, the entered SMILES notation and a list of the identified susceptible fragments and the taxa these fragments have been observed to biotransform within. The user can view this report and then has the option to save a copy of the report as a text file (by pressing the **Save** button). The summary report for the query molecule chlorfenvinphos is provided in Figure 11.



http://bioinformatics.csl.gov.uk - r.stones@csll.gov.uk

File Tools Settings Help

Biotransformation

HITID	SMILES	Atoms	No. of Reaction...
9	CO	2	26
20	CCO	3	17
293	CC	2	26
488	c1ccccc1	6	3
749	C=C	3	5
750	C#C	2	4
1264	P(O)(O)OC	5	12
1265	P(O)(OC)OC	6	12
1284	POC	3	12
1285	P(O)OC	4	12
1286	P(O)(O)OC	5	12
1322	P(O)(OC)=O	6	3
1323	P(O)(OC)(OC)=O	7	3
1330	P(O)=O	4	3
1331	P(O)(OC)=O	5	3
1332	P(O)(O)(OC)=O	6	3
1632	c1ccc(cc1)CO	8	2

Summary Results

Print Save Close

Query SMILES: CCOP(=O)(OC)OC(-CC)c1ccc(Cl)cc1Cl

Match	ray-finned fish	gastropoda	decapoda	bivalve mollusc
CO	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
CCO	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
CC	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
c1ccccc1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C=C	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
C#C	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
P(O)(O)OC	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
P(O)(OC)OC	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
POC	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
P(O)OC	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
P(O)(O)OC	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
P(O)(OC)=O	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
P(O)(OC)(OC)...	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
P(O)=O	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
P(O)(OC)=O	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
P(O)(O)(OC)=O	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c1ccc(cc1)CO	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Tora	Reactio...	Group	No. of S...	HITID
ray-fin...	Addition	glucur...	3	9
ray-fin...	Addition	S(=O)...	1	9
ray-fin...	Addition	O	1	9
ray-fin...	Reduct...	O	1	9
ray-fin...	Addition	=O	1	9
ray-fin...	Oxidati...	O	1	9
ray-fin...	Loss	O	1	9
gastro...	Loss	C	2	9
ray-fin...	Addition	C	1	9
decap...	Loss	C	3	9

Figure 11. The summary report for the query molecule chlorfenvinphos