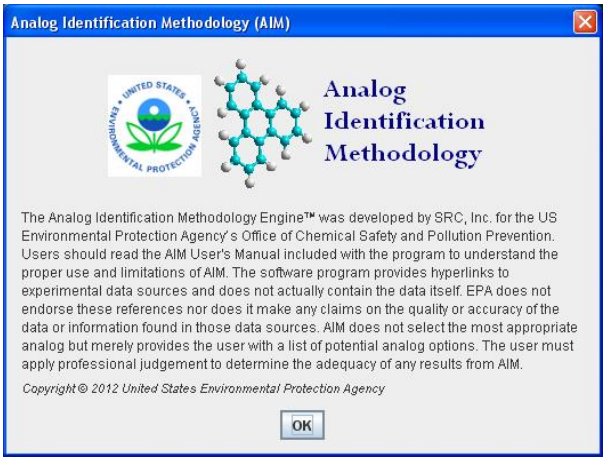
9. Analog Identification Methodology (AIM)

The Analog Identification Methodology (AIM) helps identify analogs with publicly-available measured data and when possible provides links to the data. AIM has been converted from a web application to a downloadable software program. Improvements include fragment remapping capability, and the User's Manual is incorporated into the program. No significant methodology changes were made. The AIM download page is http://www.epa.gov/oppt/sf/tools/aim.htm.



9.1 AIM Download and Set Up

AIM is a self-extracting file which is 285 MB in size once installed. It has been tested on Windows XP, Vista, and Windows 7. Once it is downloaded from the website and saved to diskette or hard drive, execute (double-click) the installation file to install the program (if you select "Run" and not "Save" at the download prompt, it should install automatically). Note: The AIM installation program includes a version of the Java Runtime Environment (JRE) which is necessary for operation of AIM. If standard set-up procedures were followed, a Desktop shortcut for AIM should be present on the user's Desktop. The program will also be available in the 'Start' menu under 'Programs'. Either double click on the desktop icon or navigate the start menu and select the AIM program from the list.

9.2 Why was AIM Developed?

AIM was developed by EPA, Office of Pollution Prevention and Toxics (OPPT), Risk Assessment Division (RAD) under Sustainable Futures to help participants in Sustainable Futures identify analogs with publicly-available measured data. As explained in chapter 8 of this document "8. Non-Cancer Human Health Hazard – Non-Cancer Screening Protocol" OPPT RAD has no computer models to predict non-cancer effects endpoints like reproductive or developmental toxicity.

Chapter 8 "Non-Cancer Human Health Hazard – Non-Cancer Screening Protocol" explains that OPPT scientists base their evaluations on experimental data for the compound of interest. If data are not available on the chemical of interest then data on *closely related analogs* are evaluated to predict

hazards. Many participants of Sustainable Futures commented that identifying chemical analogs having measured data is one of the most difficult steps of doing an assessment. OPPT RAD developed the Analog Identification Methodology (AIM) to help address this need.

The Analog Identification Methodology (AIM) will identify analogous chemicals based on the query chemical entered. The tool will also point the user to any available test data (from the sources indexed below) for both the query chemical and the analogs. Data on close analogs can then be used to inform hazard/risk judgment of chemical(s) under study.

9.3 The AIM Methodology

AIM conducts a comprehensive structural analysis of a user entered chemical using over 700 individual atoms, groups and super fragments indexed in a predefined database. It then matches them to potential analogs from a built in inventory of over 86,000 chemicals with publicly available measured data and provides links to the data sources. AIM searches may also be tailored by the user to define what types of substitutions or exclusion rules are appropriate for the search. The benefit of using the AIM methodology over other various sub-structure tools is that the analogs identified tend to be more closely related and the tool provides smaller subsets of analogs from which to choose.

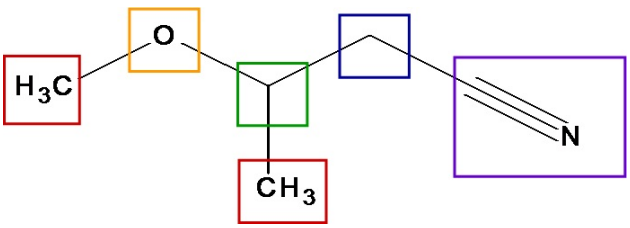
9.3.1 Online Data Sources Searched

Experimental data sources indexed by AIM include:

- Acute Exposure Guideline Levels (AEGLs)
- Agency for Toxic Substances and Disease Registry (ATSDR)
- Aggregated Computational Toxicology Resource (ACToR) Database
- Canada Chemical's Management Plan Substances (CMP) List
- Canadian Chemicals Management Plan
- Hazardous Substances Data Bank (HSDB)
- High Production Volume Information System (HPVIS)
- HPV Programs - US and OECD
- Integrated Risk Information System (IRIS)
- International Uniform Chemical Information Database (IUCLID)
- Pesticide Ecotoxicology Database
- Registry of Toxic Effects of Chemical Substances (RTECS)
- The Distributed Structure-Searchable Toxicity Database Network (DSSTox)
- The ECOTOXicology database (ECOTOX)
- The National Toxicology Program (NTP)
- The Toxic Substance Control Act Test Submission database (TSCATS)

9.3.2 Fragment-based Searching

AIM does a similarity analysis to identify analogous compounds using a chemical fragment-based approach with individual fragments indexed in the database. The image below shows sample fragments.



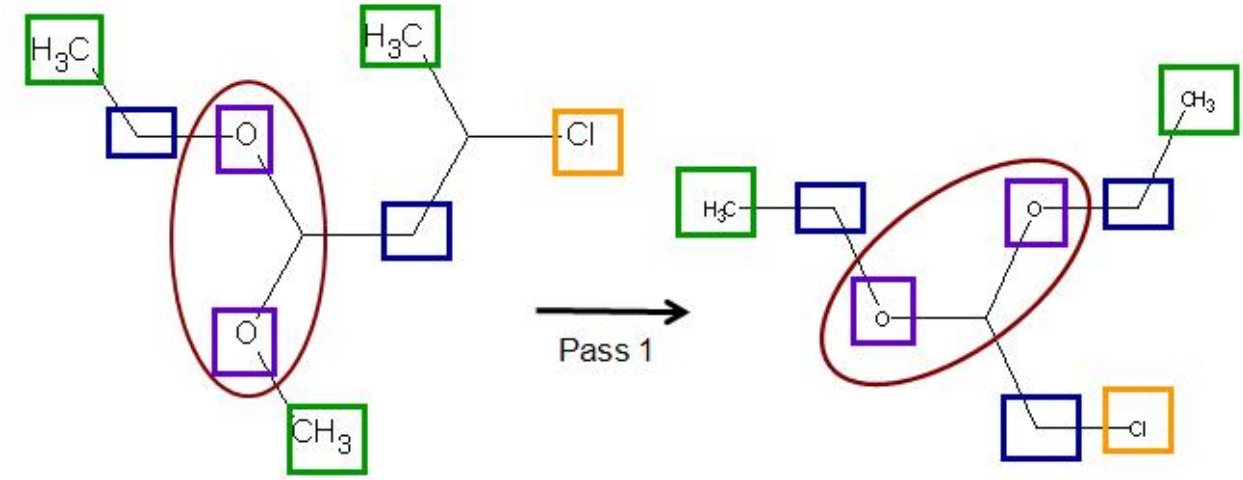
The fragments shown include:

-CH3 [aliphatic carbon]
-CH2- [aliphatic carbon]
-C#N [cyano, aliphatic attachment]
-O- [oxygen, aliphatic attach]
-CH [aliphatic carbon]

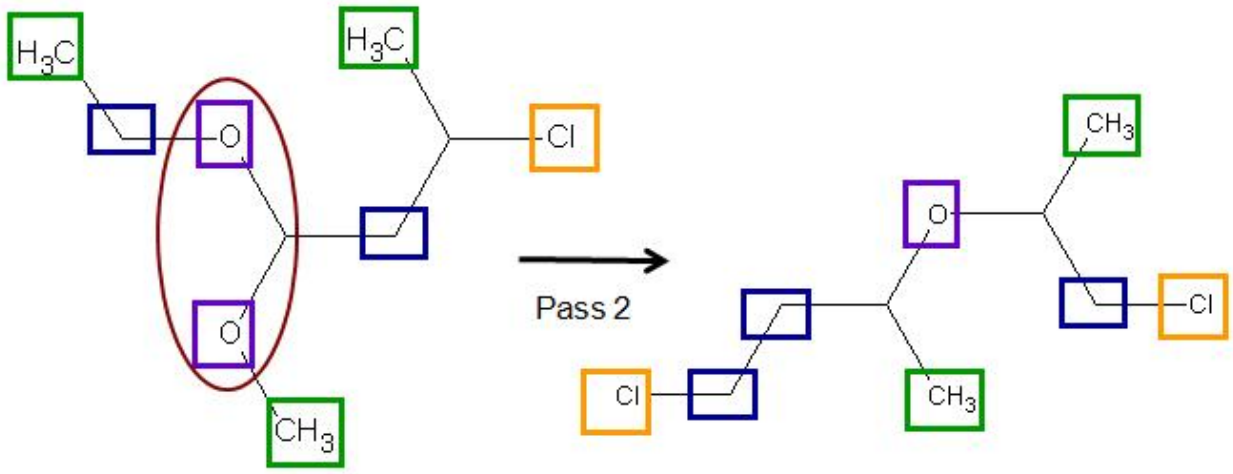
9.3.3 Pass Searching

AIM contains a tiered set of searching approaches for identifying analogs having measured data. The user can specify which passes/rules they want to use in the search.

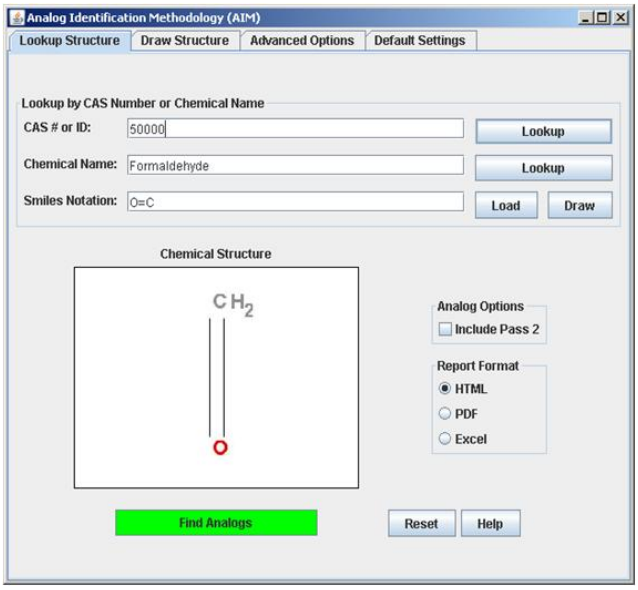
Pass 1 uses stringent search criteria. During the Default Pass 1 search chemicals are selected when all fragments/atoms/super fragments in the query chemical are contained in the analogs identified. This is the default type of search done if no other specialized rules are selected. Must be a 1:1 match!



Pass 2 – The stringent criteria from Pass 1 are loosened during the Pass 2 Selection. Many of the large super fragments that specify orientation of atoms (such as the acetal group in this example) are no longer part of the search. Other specific rules such as combining acrylates/methacrylates and allowing metal substitutions within the same period are initiated.



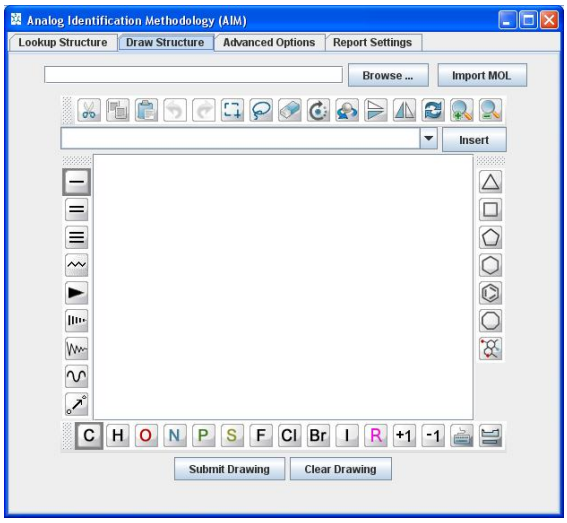
9.3.4 The Four Tabs in AIM

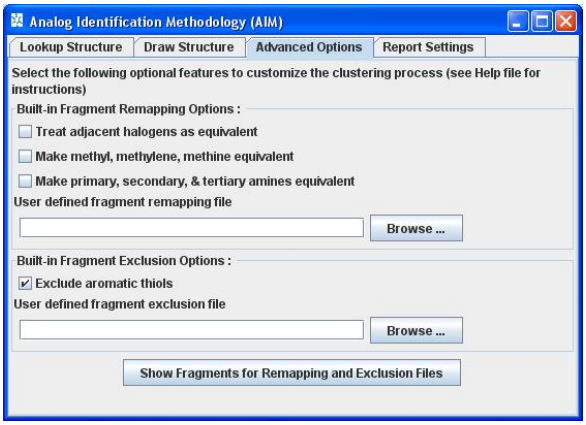


Look-up Structure Tab
When the user starts AIM, the initial screen is the "Look-up Structure" tab shown at the left. A query chemical can be entered into AIM using the CAS RN, chemical name, SMILES notation, or uploaded .mol file or a user can draw the structure with the drawing applet. The user can chose to include Pass 2 and chose the report format.

Draw Structure Tab

This page (shown on the right) allows the user to draw their chemical of interest or import the structure as a .mol file into AIM.



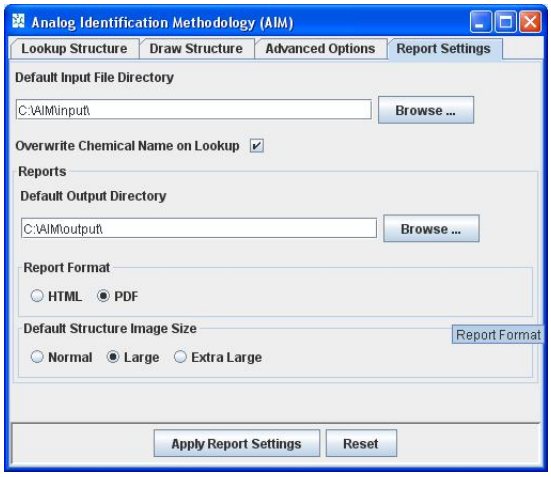


Advanced Options Tab

Check-boxes representing specialized predefined rules for advanced searching are also available to the user under the 'Advanced Options' tab (shown on the left). If a user initially finds a very limited set of analogs using only Pass 1 or 2, then these options may be used to loosen the search criteria and will often result in a larger group of identified analogs. Users can also create specific rules.

Report Settings

Under the Report Settings tab users can customize the format and file location of the output report.



9.4 Important Notes on Using AIM

Users of AIM should be aware of these important aspects of AIM:

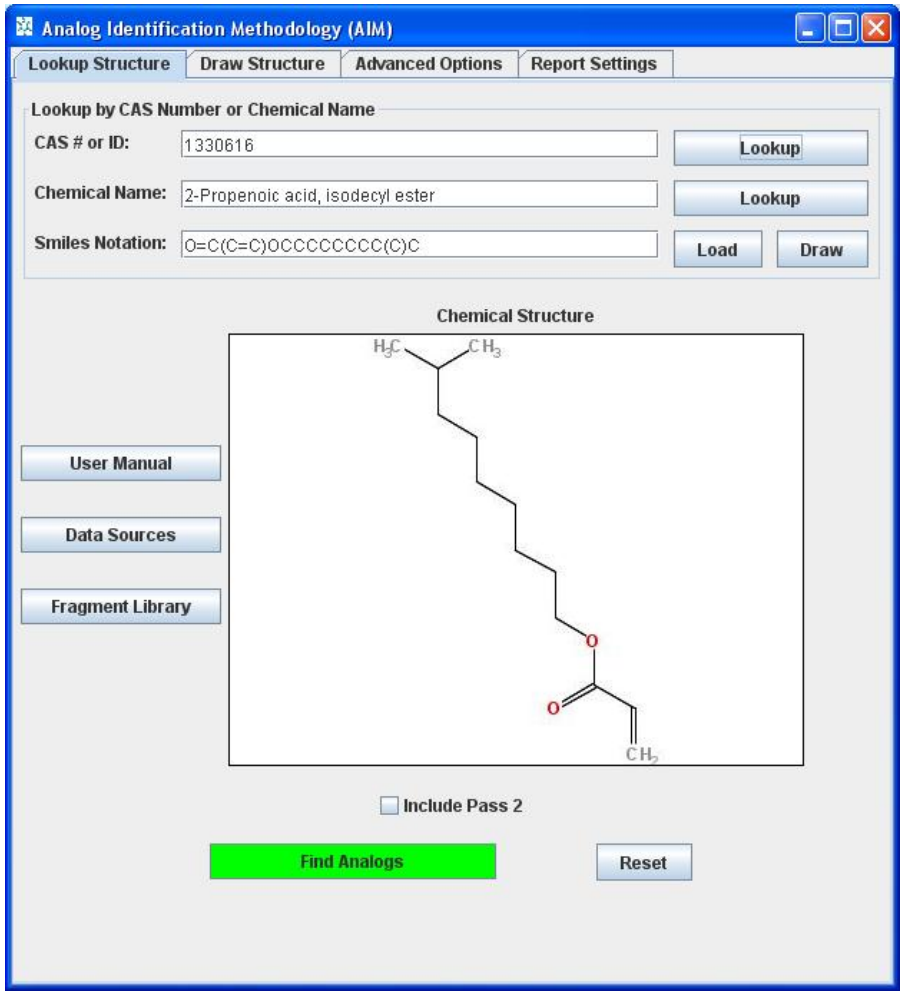
- If the chemical entered has data in the data sources being searched that chemical is listed first
- Analogs are ordered by the amount of data publicly available
- The best analogs may NOT be those that have the most data
- AIM does not rank chemical analogs from best to worst, because this varies depending on many factors

9.5 Running the Sample Chemical Isodecyl Acrylate in AIM

9.5.1 Look-up Structure Tab

When the user starts AIM, the initial screen is the "Look-up Structure" tab shown in below. A query chemical can be entered into AIM using the CAS RN, chemical name, SMILES notation, .mol file or a user can draw the structure with the drawing applet. Isocedyl acrylate can be entered in AIM using the CAS RN of 1330-61-6. This chemical is in the database incorporated in AIM and the SMILES will be retrieved after the CAS RN is entered in AIM by clicking on the Lookup button next to the CAS box

You start the analog search by clicking on the Find Analogs button.

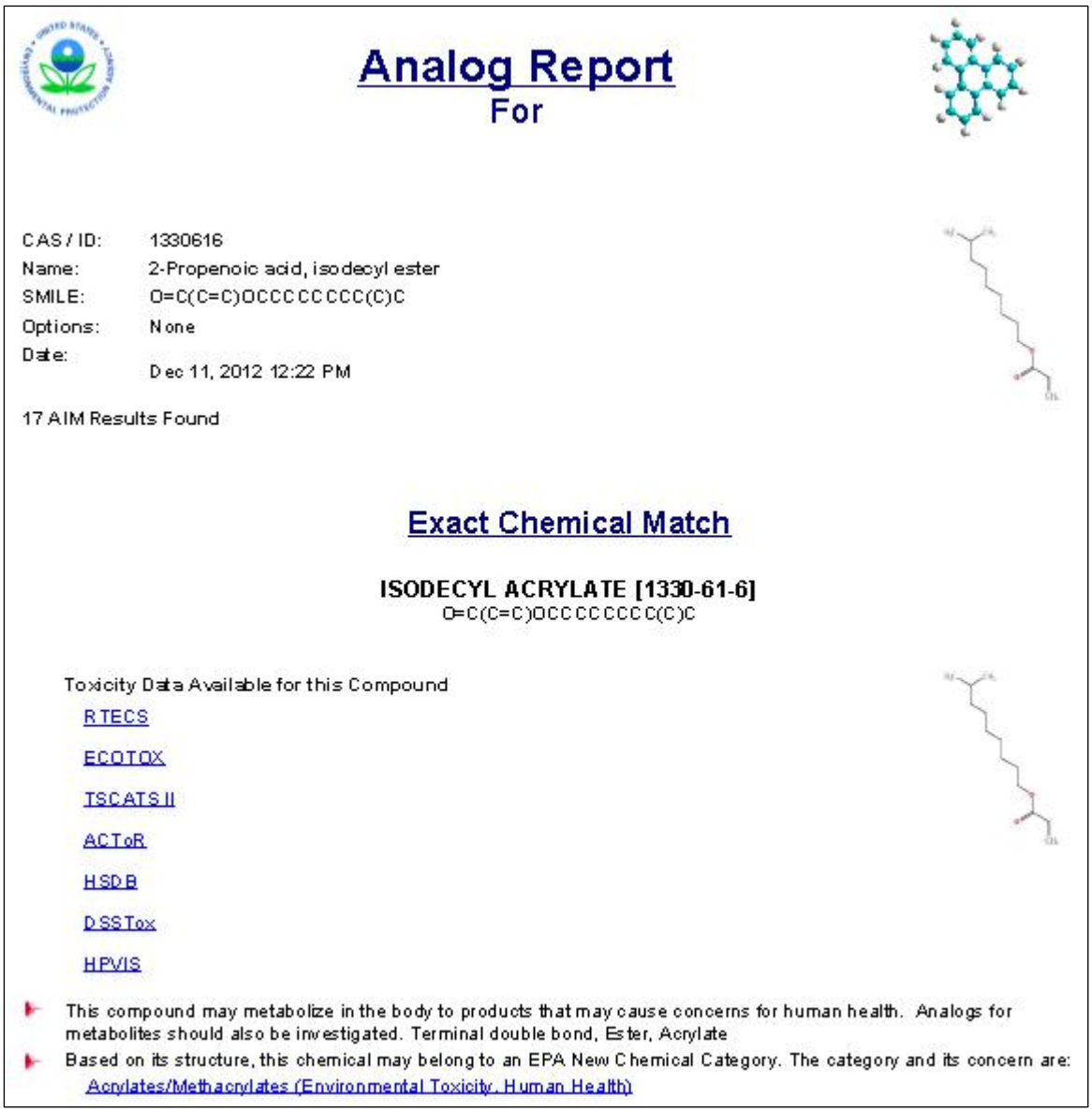


9.5.2 Results

This image below shows the results of the AIM analog search for Isocedyl acrylate. The results are presented as a PDF (or HTML). If the chemical itself is found it is identified as “Exact Chemical Match” and is displayed before any analogs will be displayed. AIM found 17 analogs for Isocedyl acrylate. Remember that analogs are presented in order by the amount of data publicly available with those having the most data presented first.

You can click on the hyperlinks for the data sources to go to the online source. If the chemical belongs to a New Chemicals Program Chemical Category you will get a link to that information.

Notice that Isocedyl acrylate triggered the metabolism flag and does belong to an EPA New Chemical Category – Acrylates.



Analog Report
For

CAS / ID: 1330616
Name: 2-Propenoic acid, isodecyl ester
SMILE: O=C(C=C)OCCCCCCCCC(C)C
Options: None
Date: Dec 11, 2012 12:22 PM

17 AIM Results Found

Exact Chemical Match

ISODECYL ACRYLATE [1330-61-6]
O=C(C=C)OCCCCCCCCC(C)C

Toxicity Data Available for this Compound

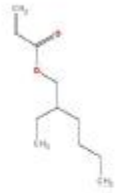
RTECS
ECOTOX
TSCATS II
ACToR
HSDB
DSSTox
HPVIS

- This compound may metabolize in the body to products that may cause concerns for human health. Analogs for metabolites should also be investigated. Terminal double bond, Ester, Acrylate
- Based on its structure, this chemical may belong to an EPA New Chemical Category. The category and its concern are:
  Acrylates/Methacrylates (Environmental Toxicity, Human Health)

This is page 2 of the search results which shows the analog having the most measured data available in the sources searched.

Analogs

2-ETHYLHEXYL ACRYLATE [103-11-7]
O=C(OCC(CCCC)CC)C=C



Toxicity Data Available for this Compound

RTECS

HPV Challenge

OECD HPV

** May also be located at:* OECD

ECOTOX

TSCATS II

ACToR

TSCATS

HSDB

IUCLID

NTP

DSSTox

HPVIS

- This compound may metabolize in the body to products that may cause concerns for human health. Analogs for metabolites should also be investigated. Terminal double bond, Ester, Acrylate
- Based on its structure, this chemical may belong to an EPA New Chemical Category. The category and its concern are:
  Acrylates/Methacrylates (Environmental Toxicity, Human Health)