

The Abstract Sifter has come a long way since its release in 2017. This document covers the full functionality of the tool.

Abstract Sifter

Version 8 User Guide



I. Abstract Sifter User Guide, Version 8.0

Availability: The Abstract Sifter and documentation are freely available for download here:

<https://www.parlezchem.com/>

Contact: Nancy Baker at nancycolebaker@gmail.com

Tony Williams at williams.antony@epa.gov

Disclaimer: The views expressed in this user guide are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

II. Contents

I.	Abstract Sifter User Guide, Version 8.0	1
A.	Introduction	4
B.	Starting out – opening the Abstract Sifter and choosing your view	4
C.	A. Main Sheet.....	7
D.	B. Abstract Sheet.....	14
E.	C. Notes Sheet.....	16
1.	Extracting and Curating data from an article	19
F.	D. Log / Batch sheet	24
G.	E. Landscape sheet.....	25
H.	F. Sample_Queries, Pathway_Queries, CuratedLists Sheets	29
I.	G. TermMap Sheet.....	32
J.	H. MeSHMine sheet	36
K.	I. Version 8 EPA custom ribbon and its capabilities.....	40
1.	Button group - Working with pdfs.....	40
2.	Button group - Get DSSTox IDs.....	42
3.	Button group – Get structure diagrams	43
4.	Button group - Chemical names, IDs, and more.....	45
5.	Button group - PubChem and Common Chemistry.....	47
6.	Button group – Formatting and File Management	49
7.	Button group - BioAssays: ToxCast and Tox21	51
8.	Button group – ToxPrint chemotypes	58
9.	Button group – Chemical lists.....	65
10.	Button group – Literature	69
L.	J. Helpful Tips and Guidelines	73
1.	Tip 1 – Checking quality of results.....	73
2.	Tip 2 – Sifting the chemical literature	73
3.	Tip 3 – Cleanup and customization	74
4.	Tip 4 – Collaborative literature review tips	75
5.	Tip 5 – Connections to the EPA Chemicals Dashboard.....	75
6.	Tip 6 - Populating the Excel Abstract Sifter from the EPA CompTox Chemicals Dashboard	79
M.	L. In-depth discussions.....	83

1.	L1. MeSH Mining in-depth.....	83
2.	L2. Run in Batch in-depth	84
N.	M. Trouble-shooting	88
1.	Extracting from pdfs	88
O.	N. Technical appendix.....	88
P.	Bibliography	92

A. Introduction

Version 8 of the Abstract Sifter features significant enhancements that make the researcher's work easier. The PubMed citation retrieval has been streamlined, ranking articles through term feature mapping have been enhanced, and the notetaking and article curation capabilities have been improved.

An emerging focus of the Abstract Sifter development is the chemical literature. The Abstract Sifter offers several ways to retrieve the chemical entities in publications and deliver them to the Abstract Sifter with the EPA unique chemical identifiers, the DSSTox identifiers (DTXIDs). Once the user has DTXIDs, the larger world of chemical data becomes accessible. Version 8 of the Abstract Sifter helps the user pull in a variety of chemical information including structure and bioactivity.

With the new features, navigation has become more complex. For this reason, many of the functions are now accessible through a custom Excel ribbon. Most of the functionality accessible through buttons and windows is still available.

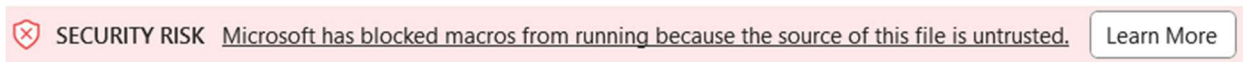
Keep in mind that the Abstract Sifter is an Excel file. You can rename it, mail it, have multiple copies, and of course, if you want to keep your Log, Notes, and Landscape entries, you should save it.

This document is roughly organized by the primary sheets in the Abstract Sifter. But we'll start this user guide with the new EPA custom ribbon.

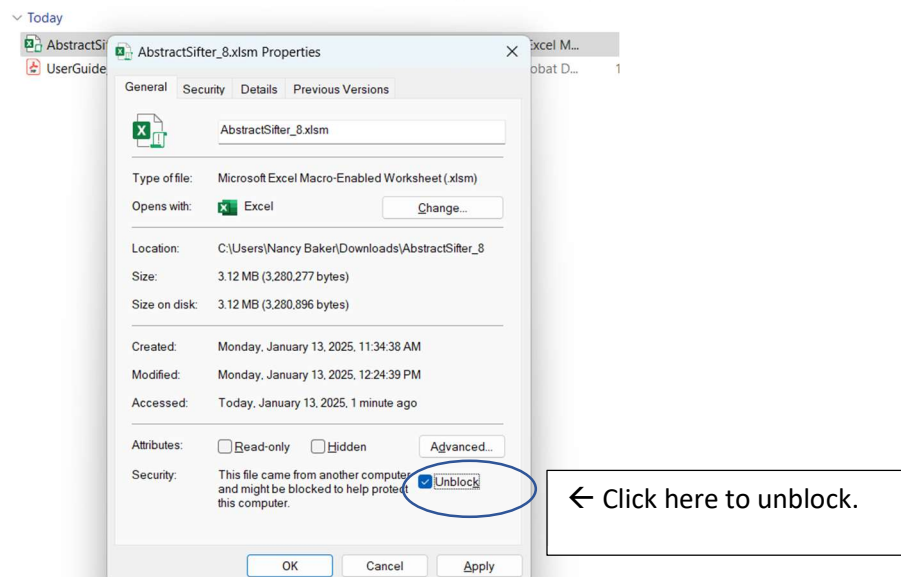
B. Starting out – opening the Abstract Sifter and choosing your view

The Abstract Sifter is an Excel workbook. You can rename it, have many copies, mail them to colleagues, and in many ways treat it just like any Excel file.

Some computers will block downloaded files. You may see this:



Try this: If the Abstract Sifter xlsx is still in a zip file, extract it. Right click on the file name and then Properties. When the properties from comes up, click on unblock. (See below)



On opening you will likely be asked to enable macros. In some environments the dialog may look different. Security varies depending on many factors that your organization has control of. There is now available a version of the Abstract Sifter that is digitally signed with a signature from a certificate authority. Contact nancycolebaker@gmail.com for this version.

The message box below may appear. Click on OK to continue.

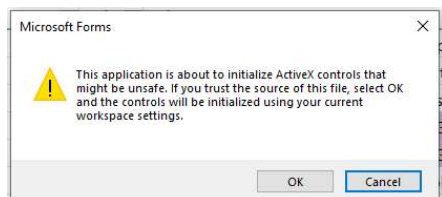


Figure B-1. This warning may appear at some point. Click on OK.

A custom Excel ribbon has been integrated into the Abstract Sifter version 8. After opening the Abstract Sifter Excel file and enabling macros, click on EPA custom to view the ribbon.



Figure B-2. The EPA custom ribbon contains controls for many new functions.

The first button on the left showing a bird sitting on binoculars and entitled *Choose view* allows the user to select the sheets (and functionality) to focus on and put out of view those not currently of interest.

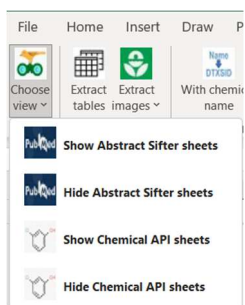


Figure B-3. Control the sheets you view.

The user can either show or hide the Abstract Sifter sheets or show or hide the (new) Chemical API sheets). See Figure B-3.

What are the Chemical API sheets? The acronym API stands for Application Program Interface and in this case, it means retrieving chemical information from the internet cloud to pull into the Abstract Sifter. Of course, the Abstract Sifter has long pulled PubMed citation information via API calls, too. The API part is not new, but the chemical sources are new to the Abstract Sifter and are relatively new to the EPA and the Center for Computational Toxicology and Exposure, the group developing these resources.

Click on the options and watch what happens to the Excel sheets.

C. Main Sheet

To see and work with the Main sheet, be sure to select Show Abstract Sifter sheets from the Choose view option on the EPA custom ribbon (Figure C-1).

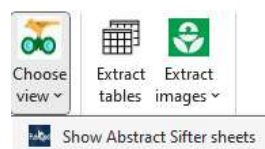


Figure C-1. Choose view with Abstract Sifter sheets.

The Main sheet is where the basic functionality of Abstract Sifter occurs, including functions we call “sifting”. To begin using the Abstract Sifter, the end-user clicks on the Query PubMed button at the top of the screen in the Main sheet. A form is displayed on which the user enters a query. In the example, we are showing a very simple query: “fipronil”, but these queries can be more complex. (Figure C-2) The text that the user enters into the box is sent to PubMed, so all PubMed syntax rules apply. (For a review of this syntax, visit <https://www.ncbi.nlm.nih.gov/books/NBK3827/>)

Note also that the end-user can optionally append the result of a query to the records currently on the Main sheet.

PMID	chorme	toxic	embryo	Score
38393619	0	15	0	15
38393234	0	0	0	0
38368720	0	4	0	4
38354522	0	2	0	2
38354521	0	6	0	6
38354108	0	1	0	1
38338357	0	1	0	1
38329667	0	1	0	1
38325757	0	1	0	1
38290626	0	2	0	2
38264609	0	3	0	3
38252984	0	1	0	1
38248665	0	1	0	1
38246053	0	6	0	6
38242394	0	1	0	1
38241915	0	1	0	1
38232547	0	1	0	1
38225075	0	3	0	3
38219570	0	4	0	4

Figure C-2. Running a PubMed query.

When finished entering the query, the user clicks on Submit and the query is packaged by a Visual Basic Application (VBA) into an e-utility command that is passed to the NCBI (National Center for Biotechnology Information) web services. (Note that using Query PubMed capability requires internet access.) The first response returned by the utility is the number of articles found. (Figure C-3) This number is displayed, and the user is asked if they want to continue. If the number of articles is over 10,000, the query will not be executed and the user is encouraged to refine the query to return fewer records. If you need to build a large corpus, see Tip 7 below on ways to do this.

The screenshot shows the EPA custom Excel interface. The 'Abstract Sifter' tool is active, and a 'Query PubMed' dialog box is open. The query entered is 'fipronil'. A 'Continue' dialog box is also open, asking 'Your query has found 1879 records. Do you wish to continue?' with 'Yes' and 'No' buttons. The background shows the 'Abstract Sifter' worksheet with columns for PMID, chlorome, toxic, embryo, and Score. The 'Abstract Sifter' tool is also visible in the top right corner.

Figure C-3. Responding to PubMed.

If the returned results are fewer than 10,000 and the user indicates the user wants to continue, the articles are downloaded from NCBI by Excel, and regular expressions are used to parse the citations for title, abstract, authors, publication year, journal, and PubMed identifier. Each record returned is inserted into a row in the Main sheet. Any rows in the Main sheet from a previous query are deleted unless the end-user chose the Append option. The Append option adds the new results to the end of the Main sheet. You can watch the status bar at the bottom of Excel to see how far along the retrieval process is. For longer queries, a cancel form will appear and the user can use it to cancel the process.

The screenshot shows the EPA custom Excel interface. The 'Abstract Sifter' tool is active, and the 'PubMed query run: fipronil' is displayed in the status bar. The 'Abstract Sifter' worksheet is visible, showing columns for PMID, chlorome, toxic, embryo, and Score. The 'Abstract Sifter' tool is also visible in the top right corner.

Figure C-4. Results from PubMed query - before sifting.

At this point the results of the query are stored in the Main sheet and can be browsed like any other data in a spreadsheet (Figure C-4); however, the most effective way to find articles of interest is to use the innovative sifter functionality. To demonstrate this functionality, we will continue to use our example of fipronil.

Let us suppose at this point that we are looking for dose-response toxicity data for fipronil. We type the term “fipronil” in cell B3, “toxic” in C3, and “mg” in D3. As we finish typing and move to the next cell, the Abstract Sifter will count the occurrences of the terms in the title, abstract, and key words combined. (The abstract and keywords are actually stored in column L which is normally hidden). The citations can then be sorted by these counts, either individually or by the total. Figure C-5 shows what the Sifter looks like when these terms have been entered into cells B3, C3, and D3 and then the entries sorted by occurrence counts of “fipronil” in descending order. PubMed article 32723848 has 28 occurrences of “fipronil”, 4 of “toxic”, and none of “mg/kg”. Looking farther down, PubMed article 12442503 has 21 of fipronil, 13 of “toxic” and four of “mg”. This article indeed describes a toxic doses of the chemical in various animal species. The sifter terms can be changed as many times as you want. Try “ppm” instead of “mg” or try “ticks”, “fleas”, “cats”, “dogs”.

	A	B	C	D	E	F	G	H	I	J
Abstract Sifter			Query PubMed							Provided by the US EPA's Center for Computational Toxicology and Exposure
1						PubMed query run: fipronil				
2	Version 8.0	Your sifter terms and frequency counts								
		fipronil	toxic	mg	Score		Take Group Notes	More things	Review	PMC
3	PMID					Pub Y	Title		Authors	
4	32723848	28	4		32	2020	Preclinical Transplacental Transfer and Pharmacokinetics of Fipronil in Rats.		Chang YN, Tsai TH	
5	31278966	23	3		26	2019	In vitro inhibition of human CYP2D6 by the chiral pesticide fipronil and its metabolite fipronil sulfone: Predict		Carrão DB, Habenchus MD, de Albuquerque NCP T	
6	30718154	23	1		24	2019	Distribution of fipronil in humans, and adverse health outcomes of in utero fipronil sulfone exposure in newb		Kim YA, Yoon YS, Kim HS, Jeon SJ, Cole E, Lee J, K I r	
7	27067106	23	10		33	2016	Fipronil sulfone induced higher cytotoxicity than fipronil in SH-SY5Y cells: Protection by antioxidants.		Romero A, Ramos E, Ares I, Castellano V, Martín T	
8	30521755	21			21	2019	Quantitative Detection of Fipronil and Fipronil-Sulfone in Sera of Black-Tailed Prairie Dogs and Rats after Or		Wang K, Vasylieva N, Wan D, Eads DA, Yang J, Tr A	
9	21615307	21			21	2011	Thyroid function tests in persons with occupational exposure to fipronil.		Herlin F, Boutet-Robinet E, Levant A, Dulaurent S T	
10	19731660	21			21	2009	Fipronil and its degradates in indoor and outdoor dust.		Mahler BJ, Van Metre PC, Wilson JT, Musgrove KE	
11	12442503	21	13	4	38	2003	Fipronil: environmental fate, ecotoxicology, and human health concerns.		Tingle CC, Rother JA, Dewhurst CF, Lauer S, King R	
12	27037470	20	5		25	2016	The toxicity, bioaccumulation, elimination, conversion of the enantiomers of fipronil in Anodonta woodiana		Qu H, Ma RX, Liu DH, Jing X, Wang F, Zhou ZQ, W Jc	
13	22447239	20	2		22	2012	CYP450-dependent biotransformation of the insecticide fipronil into fipronil sulfone can mediate fipronil-in		Roques BB, Lacroix MZ, Puel S, Gayraud V, Picard T	
14	18200855	20			20	2007	Enantioselective microbial transformation of the phenylpyrazole insecticide fipronil in anoxic sediments.		Jones WJ, Mazur CS, Kenneke JF, Garrison AW	
15	34450423	19	8		27	2021	Toxic effects of fipronil and its metabolites on PC12 cell metabolism.		Song X, Wang X, Liao G, Pan Y, Qian Y, Qiu J	
16	27614034	19	6		25	2016	Environmental behavior of the chiral insecticide fipronil: Enantioselective toxicity, distribution and transform		Qu H, Ma RX, Liu DH, Gao J, Wang F, Zhou ZQ, W V	
17	15135087	19		2	21	2004	In vitro metabolism of fipronil by human and rat cytochrome P450 and its interactions with testosterone and		Tang J, Amin Usmani K, Hodgson E, Rose RL	
18	33993965	18	2		20	2021	The effects of fipronil on emotional and cognitive behaviors in mammals.		Suzuki T, Hirai A, Khidkhan K, Nimako C, Ichise T, P	
19	22045597	18			18	2012	Adsorption, transport and degradation of fipronil termiticide in three Hawaii soils.		Shuai X, Chen J, Ray C	
20	9860498	18	7		25	1998	Mechanisms for selective toxicity of fipronil insecticide and its sulfone metabolite and desulfinyl photoprodi		Hainzl D, Cole LM, Casida JE	
21	34644892	18			18	1995	Field Efficacy of a Mechanical Pump Spray Formulation Containing 0.25% Fipronil in the Treatment and Cont		Postal JR, Jeannin PC, Consalvi PJ	
22	37201841	17	1		18	2023	Identification, occurrence, concentration and composition profile of fiproles in municipal wastewater treatm		Zhanq Q, Yang Y, Xiao Y, Xia T, Shanq N, Liu Y, Gu T	

Figure C-5. After sifter terms were entered into cells B3, C3, D3 and sorting on B3

Sifting the results through specifying sifter terms in B3, C3, and D3 can be repeated as many times as the user wishes. Similarly, new PubMed queries can be run, altered, rerun.

Let’s look at the other buttons on the Main sheet. The Take Group Notes button is discussed in section C of this guide. The More Things button contains some other actions that are often useful. Let’s check that out. Click on the More Things button reveals the menu displayed in Figure C-6.

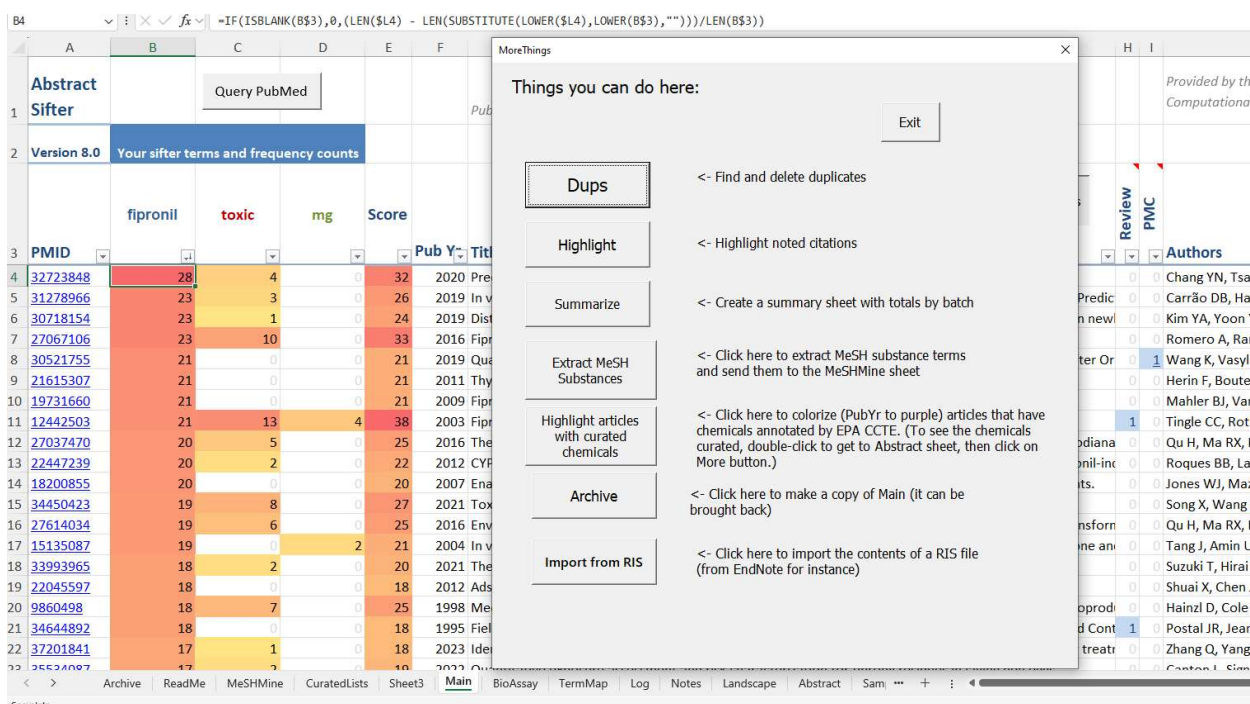


Figure C-6. The More Things button on the Main sheet

The Dups button will find and select duplicate records (duplicate records have the same PMID) and ask if you want to delete them. Let's say you run a query "fipronil AND fleas" and then run a query "fipronil AND ticks" and append the results for the second query to the first. There are bound to be duplicates. In this example, 39 duplicates were found and deleted shown in Figure C-7.

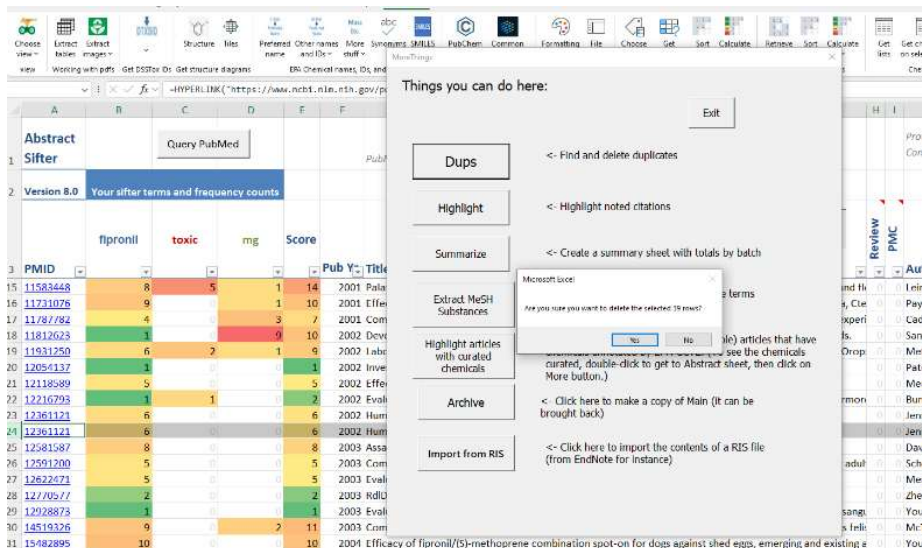


Figure C-7. Removing duplicates

The Highlight button will be covered in the Notes section of this manuscript. The Summarize function will be covered in the Log / Batch section.

Let's look at the Archive button next. Sometimes people work for hours to retrieve a set of PubMed citations through a complex set of queries. The Main sheet is then the result of a lot of work. It used to be that running another query would clear the Main sheet and some users were hesitant to do this. The Archive button will copy the records on the Main sheet to another sheet called Archive. Once that is done, the user can run any number of queries sending results to the Main sheet and then when done can go to the Archive sheet and send back the contents to Main.

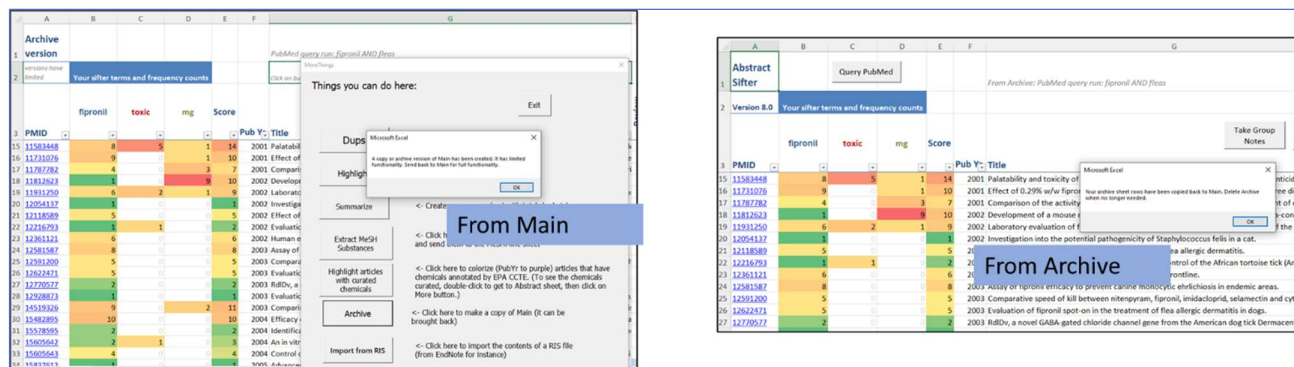


Figure C-8. Archiving and unarchiving the results on the Main sheet

Another option on the More Things menu is *Extract MeSH Substances*. See section H below for a step-by-step description.

The *Import from RIS* button allows the user to import citations in RIS format. Many citation management tools such as EndNote export to RIS format and it may be convenient to import those records. Simply click on the button and follow the instructions.

The *Highlight* button reads through the rows on the Notes sheet and colorizes the corresponding PMID on the Main sheet.

The button *Highlight articles with curated chemicals* delivers a new service under development at the EPA. To introduce this functionality, we'll need to cover some background first.

The EPA is beginning to curate chemicals from articles. The purpose is to make more chemicals findable to search engines and consequently, to give the information about those chemicals wider exposure and use. The curation project focuses on publications that measure, assay, analyze, or observe sets of chemicals. Working with sets of chemicals is becoming more common as the number of chemicals introduced into the environment increases and as methods to study them become more high-throughput. Fortunately, the technology that make it easier to extract chemical entities from the text are improving as well, and the Abstract Sifter takes advantage of those advances.

The EPA is not the only organization extracting and identifying entities in publications. The National Library of Medicine's MeSH terms for years have been the gold standard for entity extraction. But although MeSH terms are high quality, they do not reach into lists of entities in the full text of a publication. PubChem and PubTator extract from publication full text[1-3].

We use the Abstract Sifter to extract chemical terms and find their identifiers, but more importantly for the Abstract Sifter user, the publications with chemicals curated can be pulled into the Abstract Sifter and the chemicals annotated can be retrieved.

Let's start with the *Query PubMed* button on the Main sheet. (Figure C-9.) From the Main sheet, identifying publications with curated chemicals can be done in two ways. First click on *Query PubMed*, then at the bottom of the form, click on *Advanced*, then on a new form click on the only button.

This action retrieves the PubMed information on the set of articles for which EPA has a set of curated chemicals and writes it to the Main sheet.

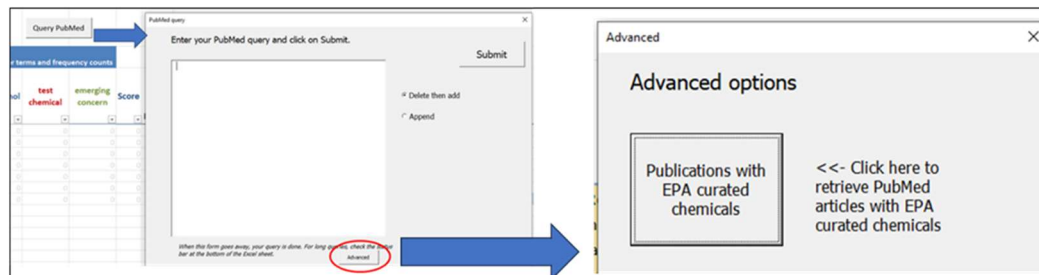


Figure C-9. How to get send the EPA curated chemicals to the Main sheet

The entries can be browsed and sifted. To see the chemicals which have been annotated, double-click on a row to view the article in the Abstract sheet. Then click on the EPA custom ribbon and on *Publication chemical retrieval* button, select *Get EPA CCTE curated chemicals*. The chemicals will be retrieved and written below the abstract with DSSTox identifies and preferred chemical names. Note that if you want to keep the list of chemicals, they should be copied to another sheet (CuratedLists for instance). The abstract sheet is cleared when new abstracts are displayed.

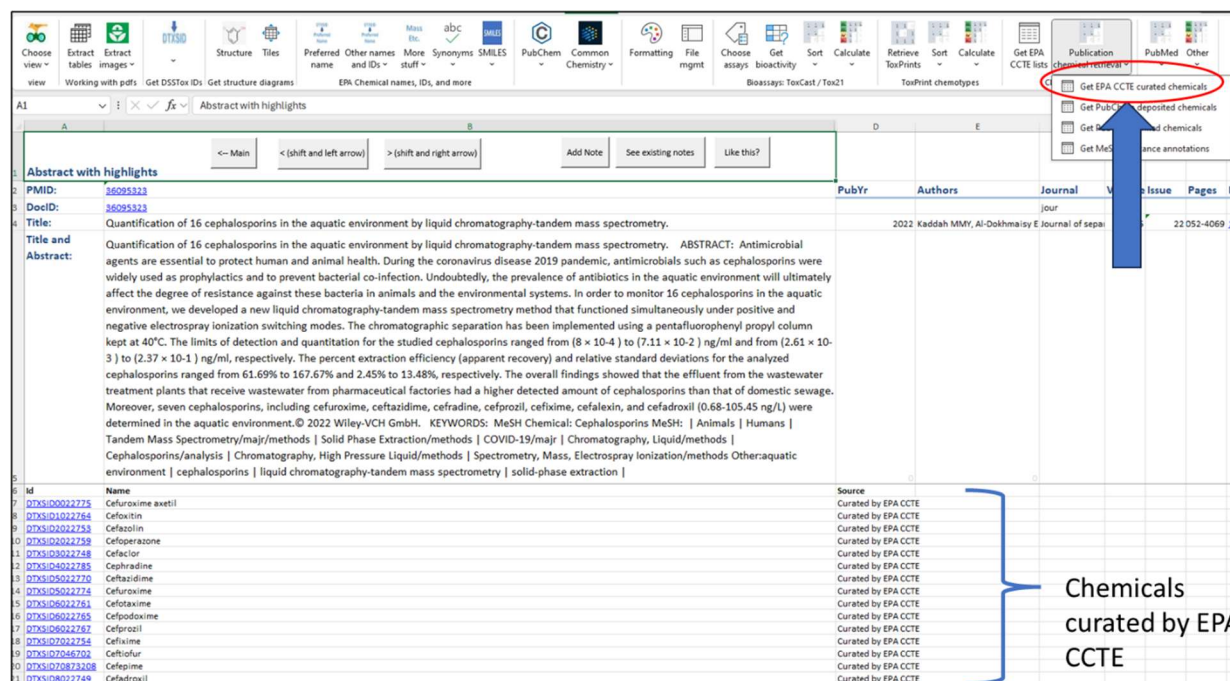


Figure C-10. On Abstract sheet - how to retrieve curated chemicals.

See Section I and Button group – Chemical lists for more information on this functionality.

Often users will run queries, retrieve articles, and then wonder if any of the articles on the Main sheet have EPA curated chemicals. This is where the button on the *More things* form comes in handy.

Let's say that we run this query: "embryonic development AND stem cell AND assay AND toxicity". We wonder if any of the 747 articles returned has a chemical list curated that would be useful in designing a new assay. Click on *More things*, then *Highlight articles with curated chemicals*. The Abstract Sifter will tell you how many articles have annotated chemicals and then color the publication year (Pub Yr) cell purple. The illustration here has the *Pub Yr* column sorted by color.

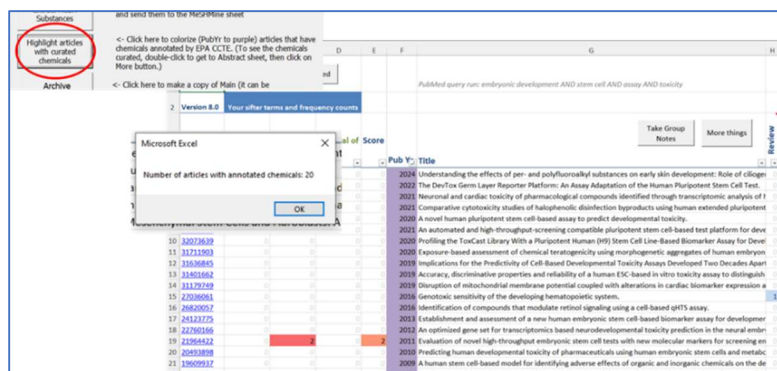


Figure C-11. How to highlight PubMed articles with chemical lists curated by the EPA.

D. Abstract Sheet

To see the abstract for any of the retrieved articles, the user can either click on the PubMed ID hyperlink to be taken to PubMed, or double-click on any cells in the row except for the PubMed ID link in Column A. This action brings the user to the Abstract sheet where the abstract is displayed along with other article meta-data such as title and authors (Figure D-1).

A		B			D	E	F
1	Abstract with highlights	← Main < (shift and left arrow) > (shift and right arrow) Add Note See existing notes Like this?					
2	PMID:	32723848			PubYr	Authors	Journal
3	DocID:	32723848					jour
4	Title:	Preclinical Transplacental Transfer and Pharmacokinetics of Fipronil in Rats.			2020	Chang YN, Tsai TH	Drug metab
5	Title and Abstract:	<p>Preclinical Transplacental Transfer and Pharmacokinetics of Fipronil in Rats. ABSTRACT: Fipronil, a widely used insecticide and pesticide, with its toxic metabolite fipronil sulfone was detected in fipronil-contaminated eggs as a result of inappropriate use. However, little is known about whether fipronil and fipronil sulfone transfer into fetus through the blood-placenta barrier. Our objectives were to investigate the transplacental transfer and the pharmacokinetics of fipronil and fipronil sulfone in rats. Male and female (with 13 days of gestation) Sprague-Dawley rats were used in pharmacokinetics and transplacental transfer experiments, respectively. Biologic samples were collected at each time point after fipronil intravenous or oral administration. To monitor fipronil and fipronil sulfone in the plasma, placenta, amniotic fluid, and fetus, a validated liquid chromatography tandem mass spectrometry method was developed. After fipronil administration in male rats, the oral bioavailability decreased, whereas the biotransformation increased as the dose increased, revealing an enhancement of first-pass effect and a fast metabolism in vivo. The results of fipronil transplacental transfer in pregnant rats demonstrated that the concentration of fipronil and fipronil sulfone varied in the following order, respectively: placenta > plasma > fetus > amniotic fluid and plasma > placenta > fetus > amniotic fluid. This is the first direct evidence that fipronil and fipronil sulfone cross the blood placental barriers and enter the fetus. The amount of fipronil distributed to the fetus was greater than that of fipronil sulfone in the short term, but by contrast, pharmacokinetic data showed that the latter stayed longer in the body. These findings provide constructive information for public health alarm. SIGNIFICANCE STATEMENT: Fipronil and fipronil sulfone interfere with the GABAergic system. Fipronil can cause thyroid dysfunction, which may affect brain growth and nerve development. Although we knew that fipronil and fipronil sulfone could enter eggs, there was no direct evidence that they would enter fetuses. This research provided evidence on the pharmacokinetics and transplacental transfer of fipronil and fipronil sulfone, confirming our hypothesis. Copyright © 2020 by The American Society for Pharmacology and Experimental Therapeutics. KEYWORDS: MeSH Chemical: Insecticides Pyrazoles fipronil sulfone fipronil MeSH: Administration, Oral Amniotic Fluid/chemistry Animals Biological Availability Biotransformation Dose-Response Relationship, Drug Female Fetal Blood/majr/chemistry Insecticides/majr/pharmacokinetics/toxicity Male Maternal-Fetal Exchange/majr Placental Circulation Pregnancy Pyrazoles/majr/analysis/pharmacokinetics/toxicity Rats Toxicity Tests, Acute</p>					
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							

Figure D-1. An abstract with highlighted sifter terms. Fipronil is blue, toxic is red, and mg is green.

There are several aspects of the Abstract sheet that are important to note. First, the sifter terms in the abstract are highlighted. The font colors reflect the colors of the fonts in cells B3, C3, and D3 entered into the Main Sheet. This highlighting makes reading the abstract easier by drawing attention to sentences that might be of more interest.

Let's go through the other buttons on the Abstract sheet. There are three navigation buttons. One takes the user back to the Main sheet. The next two left/right navigation buttons retrieve the previous or next row on the Main sheet. The user can either click on the buttons or use the keyboard shift and left or right arrows.

The *Like this?* button allows the user to find articles in PubMed that have a relationship to the article on the Abstract sheet. PubMed provides this information for retrieval. Let's take a look at PubMed ID 30521755 to see what kind of articles are made available from PubMed Entrez web site. PubMed offers Similar articles, Cited by, and References. Not all of these options are available for all articles. See Figure D-2 and here for more information:

<https://pubmed.ncbi.nlm.nih.gov/help/#similar-articles>

Figure D-2. Clicking on the *Like this?* button on the Abstract Sheet allows retrieval of article lists provided by PubMed.

Figure D-3. Choosing connected PubMed citations

The results can be appended to results already on the Main sheet. In addition, the title for the added articles can be colored green or yellow to help distinguish them from the previous results on the Main sheet. This way the user can see which rows were from the original query and which were retrieved from the *Like this* functionality. (Figure D-3)

If you have already taken a Note on this article you can click on the See Existing Notes button to be taken to that citation on the Notes sheet. If no note is yet taken, a message saying that will be displayed. The Add Note button brings up the Notes form and you can enter a Note.

E. Notes Sheet

Given the dynamic nature of the Abstract Sifter, many users find it helpful to be able to make permanent notes on articles that they want to keep track of. (They are only permanent if the Abstract Sifter is saved!)

There are two ways using the Sifter to take notes: one way is through the Main sheet, and the other way starts with the Abstract sheet. To return to our fipronil case study, let us say that we have found a set of articles on the Main sheet that we know we need to read in depth. We can select these articles and then click on the *Take Group Notes* button. A form appears where we can enter information into fields called Tag and Notes. (Figure E-1) These elements are self-defined. We can also click on yes, no, or maybe. The note-taker can enter her/his initials or name in the Who field. This set of variables is a quick way to associate articles with a note. Notice that these choices each come with a color (yes-green, no-red, and maybe-yellow). Entering any of these fields is optional. (Figures E-1 and E-2) When we click on the OK button, each article selected will be inserted into the Notes page with the corresponding information (Figure E-3) and the PubMed ID (PMID) on the Main sheet will be colored.

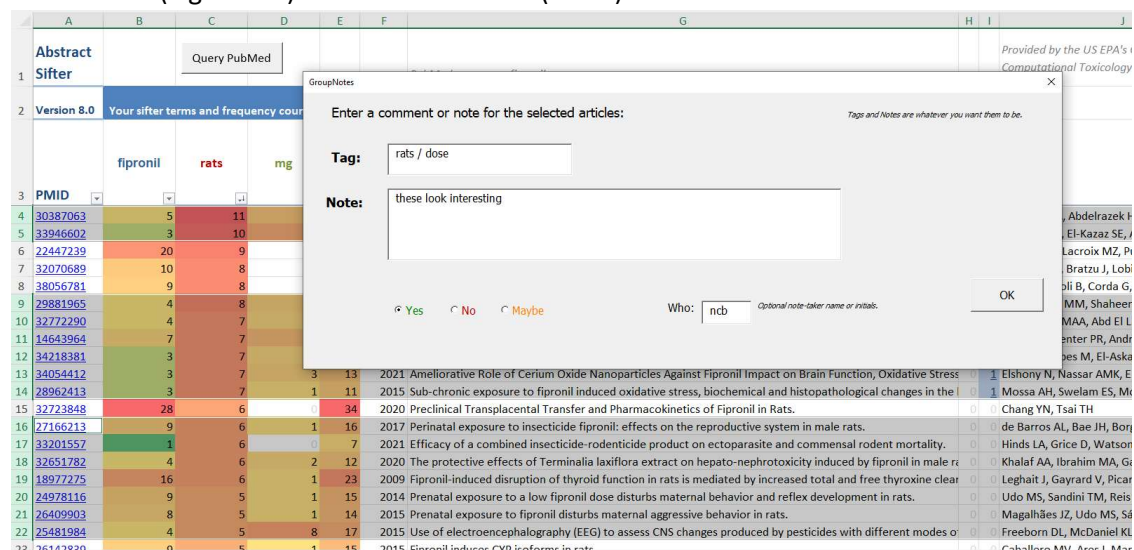


Figure E-1. Taking group notes on the Main sheet

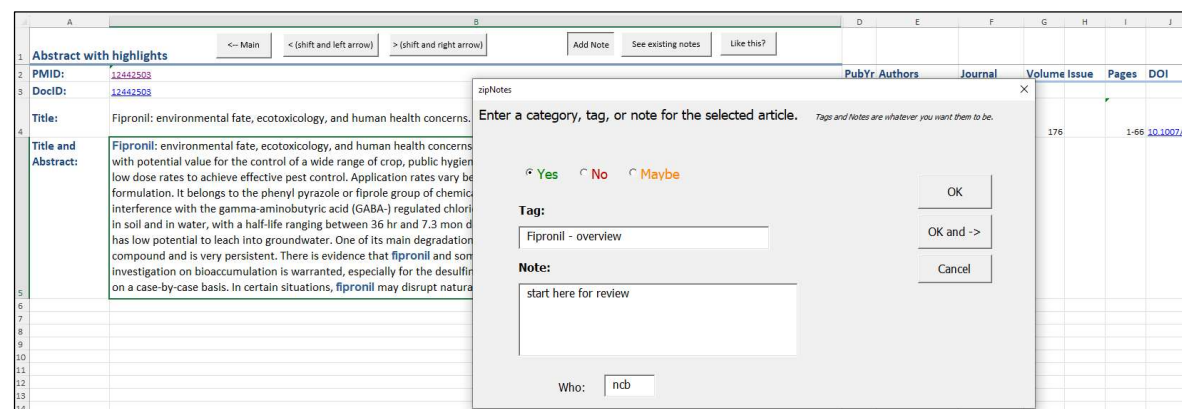


Figure E-2. Taking single notes on the Abstract sheet

	A	B	C	D	E	F	G	H	I	
1	My Notes						<- Main	Highlight Noted PMIDs	More stuff	Note: Feel free to add rows after 10 columns. So Notes may be added.
2	PMID	Yes	No	Maybe	Who	Tag	Notes	PubY	Title	Authors
3	30387063	1	0	0	ncb	rats / dose	these look interesting		2019 Hepatoprotective activity of Uncaria tomentosa extract against sub-chronic exposure to fipronil in male rats.	Elgawish RA
4	33946602	1	0	0	ncb	rats / dose	these look interesting		2021 Ameliorative Effects of Boswellic Acid on Fipronil-Induced Toxicity: Antioxidant State, Apoptotic Markers, and Testicular Steroidogenesis	Tohamy HG
5	29881965	1	0	0	ncb	rats / dose	these look interesting		2018 Thymoquinone and diallyl sulfide protect against fipronil-induced oxidative injury in rats.	Abdel-Daim
6	32772290	1	0	0	ncb	rats / dose	these look interesting		2020 Ginseng attenuates fipronil-induced hepatorenal toxicity via its antioxidant, anti-apoptotic, and anti-inflammatory activities in rats.	Abd Eldaim
7	14643964	1	0	0	ncb	rats / dose	these look interesting		2004 Reproductive adverse effects of fipronil in Wistar rats.	Ohi M, Dals
8	34218381	1	0	0	ncb	rats / dose	these look interesting		2021 Ephedra sinica mitigates hepatic oxidative stress and inflammation via suppressing the TLR4/MyD88/NF-κB pathway in fipronil-treated rats.	Seif M, Deal
9	34054412	1	0	0	ncb	rats / dose	these look interesting		2021 Ameliorative Role of Cerium Oxide Nanoparticles Against Fipronil Impact on Brain Function, Oxidative Stress, and Apoptotic Cascades	Elshony N, I
10	28962413	1	0	0	ncb	rats / dose	these look interesting		2015 Sub-chronic exposure to fipronil induced oxidative stress, biochemical and histopathological changes in the liver and kidney of male rats.	al Mossa AH, S
11	27166213	1	0	0	ncb	rats / dose	these look interesting		2017 Perinatal exposure to insecticide fipronil: effects on the reproductive system in male rats.	de Barros A
12	33201557	1	0	0	ncb	rats / dose	these look interesting		2021 Efficacy of a combined insecticide-rodenticide product on ectoparasite and commensal rodent mortality.	Hinds LA, Gr
13	32651782	1	0	0	ncb	rats / dose	these look interesting		2020 The protective effects of Terminalia laxiflora extract on hepato-nephrotoxicity induced by fipronil in male rats.	Khalaf AA, Il
14	18977275	1	0	0	ncb	rats / dose	these look interesting		2009 Fipronil-induced disruption of thyroid function in rats is mediated by increased total and free thyroxine clearances concomitantly to increased thyroidal iodine uptake.	Leghait J, Gi
15	24978116	1	0	0	ncb	rats / dose	these look interesting		2014 Prenatal exposure to a low fipronil dose disturbs maternal behavior and reflex development in rats.	Udo MS, Sa
16	26409903	1	0	0	ncb	rats / dose	these look interesting		2015 Prenatal exposure to fipronil disturbs maternal aggressive behavior in rats.	Magalhães
17	25481984	1	0	0	ncb	rats / dose	these look interesting		2015 Use of electroencephalography (EEG) to assess CNS changes produced by pesticides with different modes of action: effects of permethrin and fipronil.	Freeborn DI
18	12442503	1	0	0	ncb	Fipronil - overview	start here for review		2003 Fipronil: environmental fate, ecotoxicology, and human health concerns.	Tingle CC, R

Figure E-3. The Notes sheet. Be sure to save your workbook.

The second option for note taking starts with the Abstract Sheet. (Figure E-2) The “Add Note” button in the top row allows notes to be inserted into the Notes Sheet.

The note-taking can be used to help keep track of which citations have been read and evaluated and which have not. On the Main sheet the PMIDs can be sorted by the noted color using the built-in Excel sorting functionality.

The user can make changes to the Notes sheet by editing, adding, or deleting rows below row 2.

Double-clicking on a row in the Notes sheet brings up a curation form (Figure E-4) that makes reading and note-taking easier. Text in the large box can be selected and dragged to any of the smaller tag and notes boxes. The Refresh button brings the title and abstract back in and refreshes the view. The copy

button sends the elements of the citation and your notes to the clipboard; then you can open up Word and paste the clipboard contents there.

Figure E-4. Double-click on a row on the Notes sheet to curate your notes about the article

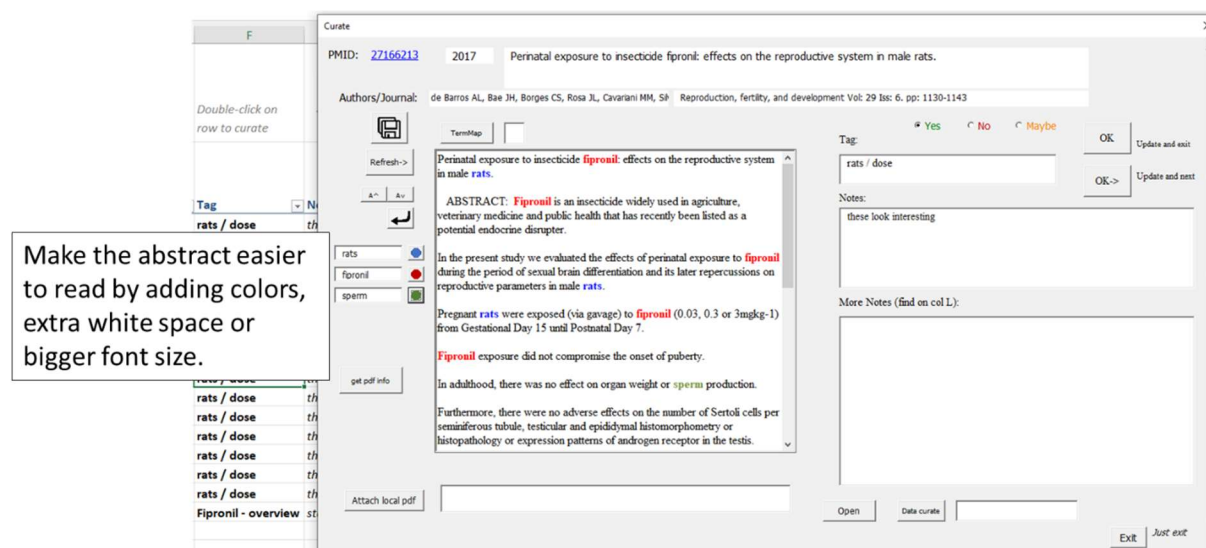


Figure E-5. The notes curation form has lots of ways to make the abstract easier to read and understand.

In version 8, new functionality to help make the abstract more readable has been added. In this example, a carriage return has been added to each sentence end to increase white space, the font size has been increased, and selected terms colorized. (Figure E-5)

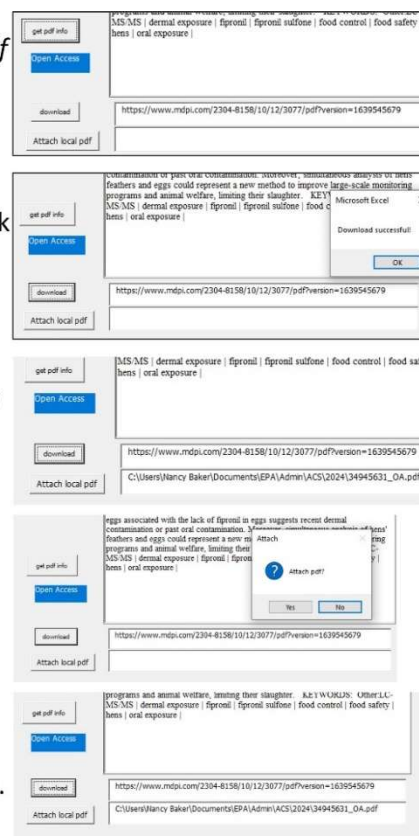
Click on the *get pdf info* button

If open access, click on download

You will be asked to select a directory

Attach the pdf to store the link

Be sure to click on OK to save the links.



Working with PDF documents is easier, too. Clicking on the *get pdf info* button sends an API call to the unpaywall service. (For more information, see <https://unpaywall.org/>) The returned information is parsed to see if there is an open access PDF available. Let's walk through the steps to find and download an open access pdf. First, click on the *get pdf info* button. (Figure E-6)

The user can also attach a pdf already downloaded and stored on the computer. Click the *Attach Local PDF* button and find the pdf. You can open the pdf by clicking on the *Open* button or on the pdf hyperlink in column W. Note that while Abstract Sifter files can be shared with colleagues, the PDFs and their links are local and specific to one machine.

Figure E-6. Series of steps to get PDF

Extracting and Curating data from an article

The new version of the Abstract Sifter includes powerful features to extract information and data from PDFs, Word, or other Excel files. This activity is called Data Curation and it can be initiated by clicking on the *Data curate* button on the Curate form. (Note: you can also just create a blank sheet and do most of the same activities.)

The Data Curate button when clicked will create a new sheet named with the PubMed ID and the first few letters of the first author's name. The PMID, title, and a link to the pdf if attached to the Curate form will be placed at the top of the sheet. You can think of this sheet as where you would put any data or images of interest to your project or your exploration of the literature.

The examples that will be shown here are relevant to our work at the EPA so they revolve around chemicals and information around chemicals. The methods can just as easily be applied to any domain – genetic research, disease research, whatever. This documentation will focus on chemicals.

Another important observation to emphasize is that PDF files vary greatly in their formats and ease of information extraction. Some PDFs have simple, clean tables that can be extracted easily and others have tables that thwart automatic methods to read. Some authors do not put their studied chemicals in a table, but instead list them in the body of the article. Many authors use abbreviations for chemical names.

Let's walk through an example.

Let's start with PubMed ID 32271623. <https://pubmed.ncbi.nlm.nih.gov/32271623/>

This article is about testing perfluorinated compounds in zebrafish. A quick way to get to the curation sheet is to go to the Main sheet, click on Query PubMed, then enter 32271623[uid].

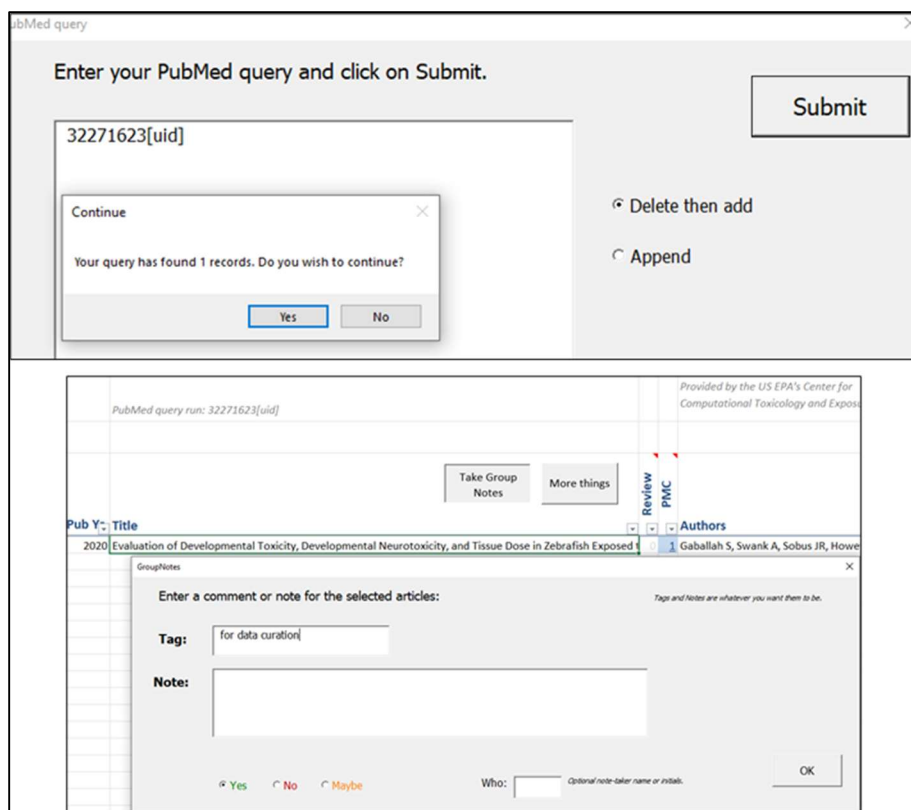


Figure E-7. Retrieving the example article.

When the article is returned to the Main sheet, select the row and click on Take Group Notes, then OK. This simple action sends the record to the Notes sheet.

On the Notes sheet, double-click on the row for the article. On the curation form, click on *Get pdf info*. The message returned says that the pdf is open access, so then click on *Download*. When asked, say yes to the request to attach the pdf. Then click on the *Data curate* button. A new sheet is inserted into the workbook with the name constructed from the PubMed ID and author name.

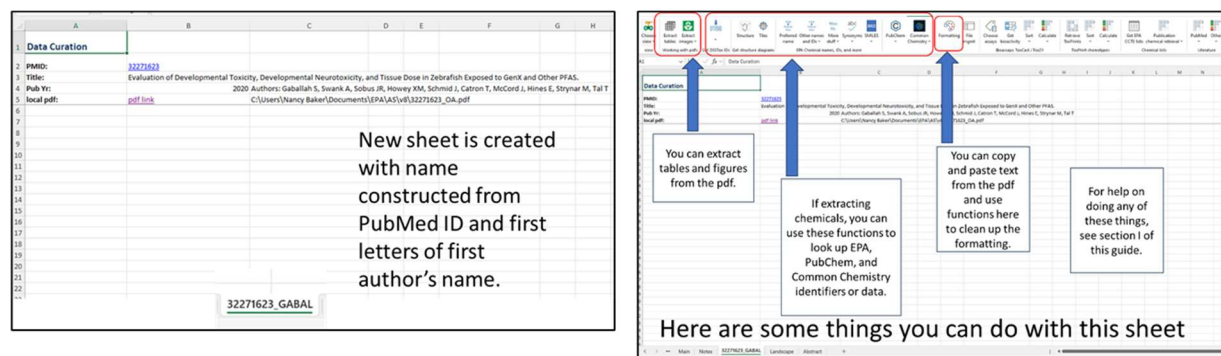


Figure E-8. Curation sheet for one article and some of the buttons that will be useful.

In this example, the tables were extracted from the pdf, the table of interest was selected and reformatted. The DSSTox ID was looked up for each chemical and the InChI key retrieved.

A detailed explanation of each of these steps can be found in Section K below.

After pdf table extraction

Data Curation

PMID: 32271623
 Title: Evaluation of Developmental Toxicity, Developmental Neurotoxicity, and Tissue Dose in Zebrafish Exposed to GenX and Other PFAS.
 Pub Yr: 2020 Authors: Gaballah S, Swank A, Sobus JR, Howey XM, Schmid J, Catron T, McCord J, Hines E, Strynar M, Tal T
 local pdf: pdf link C:\Users\Nancy Baker\Documents\EPA\AS\vs\32271623_OA.pdf

Table object (not nested) but likely not a table ... 1 row only

*These authors contributed equally to this work. Address correspondence to Tamara Tal, Department of Bioanalytical Ecotoxicology, Helmholtz Centre for Environmental Research-UFZ, Permosers

copy and paste title here from pdf and check nested results around page 3

Chemical	Name	CASRN	MW (g/mol)	LogP	Company, catalog no.
4,8-Dioxo-3H-perfluorononanoate	ADONA	958445-44-8	400.05	3.96	Wellington Laboratories, NaNODA
Perfluoro-2-propoxypropanoic acid	GenX Free Acid	13252-13-6	330.05	3.21	Synquest, 2121-3-13
Perfluorobutanesulfonic acid	PFBS	375-73-5	300.1	3.10	Synquest, 6164-3-09
Perfluoro-3,6-dioxo-4-methyl-7-oxotene-1-sulfonic acid	PFESA1	29311-47-9	444.12	6.02	Obtained from Chemours
Perfluoroheptanesulfonic acid	PFHxS	375-92-8	450.12	2.83	Synquest, 6164-3-25
Perfluorohexanoic acid	PFHxA	307-24-4	314.05	2.78	Synquest, 2121-3-39
Perfluorohexanesulfonic acid	PFHxS	3871-99-6	438.21	3.87	Synquest, 6164-3-X4
Perfluoro-n-octanoic acid	PFOA	335-67-1	414.07	3.79	Synquest, 2121-3-18
Perfluorooctanesulfonic acid	PFOA	1763-23-1	500.13	2.77	Synquest, 6164-3-08
Perfluoropentanesulfonic acid	PFPeS	2706-91-4	350.11	3.18	Synquest, 6164-3-2U

Table object (not nested) but likely not a table ... 1 row only

/Figure 1. Study design. Zebrafish were semi-statically exposed to test PFAS daily, from 0-5 dpf. At 6 dpf, developmental toxicity, developmental neurotoxicity, and PFAS tissue concentrations were

After retrieval of DTXSID, preferred name and InChI key

Data Curation

PMID: 32271623
 Title: Evaluation of Developmental Toxicity, Developmental Neurotoxicity, and Tissue Dose in Zebrafish Exposed to GenX and Other PFAS.
 Pub Yr: 2020 Authors: Gaballah S, Swank A, Sobus JR, Howey XM, Schmid J, Catron T, McCord J, Hines E, Strynar M, Tal T
 local pdf: pdf link C:\Users\Nancy Baker\Documents\EPA\AS\vs\32271623_OA.pdf

DTXSID	EPA Preferred name	Chemical name from publication	CASRN	InChI Key
DTXSID000874026	Ammonium 4,8-dioxo-3H-perfluorononanoate	4,8-Dioxo-3H-perfluorononanoate	958445-44-8	PVWXPSPVCHPTT-UMFFFAOYSA-N
DTXSID00080215	Perfluoro-2-methyl-3-oxahexanoic acid	Perfluoro-2-propoxypropanoic acid	13252-13-6	CSBNABAWMZWIF-UMFFFAOYSA-N
DTXSID00080000	Perfluorobutanesulfonic acid	Perfluorobutanesulfonic acid	375-73-5	QSTNAGHABQOMC-UMFFFAOYSA-N
DTXSID000892354	Perfluoro-3,6-dioxo-4-methyl-7-oxotene-1-sulfonic acid	Perfluoro-3,6-dioxo-4-methyl-7-oxotene-1-sulfonic acid	29311-47-9	AUJAGQGESEHBO-UMFFFAOYSA-N
DTXSID00089920	Perfluoroheptanesulfonic acid	Perfluoroheptanesulfonic acid	375-92-8	QHSQVQRYKATEL-UMFFFAOYSA-N
DTXSID00031862	Perfluorohexanoic acid	Perfluorohexanoic acid	307-24-4	PXUJGAPFEXVAM-UMFFFAOYSA-N
DTXSID00037239	Potassium perfluorohexanesulfonate	Perfluorohexanesulfonic acid	3871-99-6	QSCQGRFYSWQIF-UMFFFAOYSA-N
DTXSID00031860	Perfluoro-n-octanoic acid	Perfluoro-n-octanoic acid	335-67-1	SNQRIJHAYWQDS-UMFFFAOYSA-N
DTXSID00031864	Perfluorooctanesulfonic acid	Perfluorooctanesulfonic acid	1763-23-1	YPSUTAPUNCNZ-UMFFFAOYSA-N
DTXSID00082600	Perfluoropentanesulfonic acid	Perfluoropentanesulfonic acid	2706-91-4	AGJQXKXZQWQNT-UMFFFAOYSA-N

Figure E-9. Top figure is what the data looks like after PDF table extraction. The bottom has been formatted and InChI key added. Retrieving the DTXSID using CASRN worked better than retrieving using chemical name from publication.

The More stuff button on the top of the Notes sheet offers a number of features that might prove useful. (Figure E-10) Let's go through them.

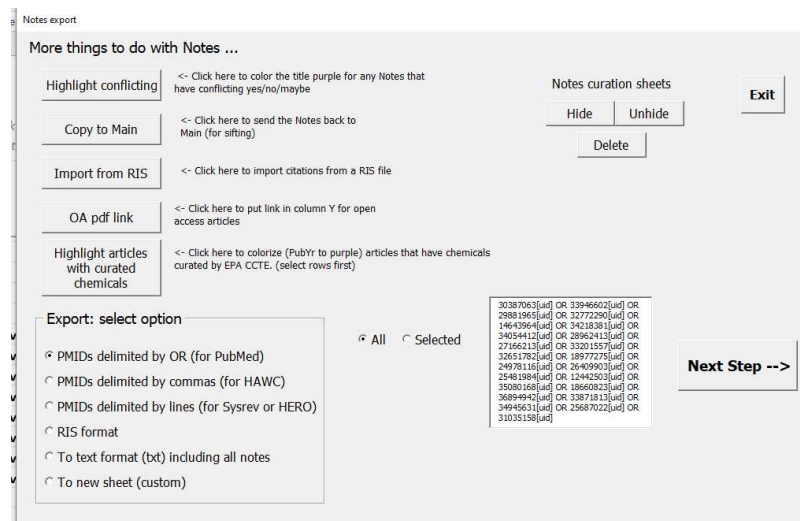


Figure E-10. The More stuff button expands what you can do with notes.

First the Highlight conflicting button will read through all the notes on the Notes sheet and if it finds two notes for the same PubMed citation that have different yes/no/maybe designations, it will color the title purple. Why would someone need such a feature? It is possible to take multiple notes on a citation. There's actually no limit. People collaborating on a project and evaluating the same corpus can have differing opinions and may designate articles differently. If it's

important to find this out and discuss the differences, the notes from each person can be combined on a Notes sheet and discrepancies can be identified and hashed out. In the example here, one can readily see the articles that Joe and Mary agree and disagree on. You can also combine notes from collaborators by copying and pasting rows from one Abstract Sifter to another.

				Who	Tag	Notes	PubY	Title
	1	0	0	mary	Fipronil - overview	good info	2003	Fipronil: environmental fate, ecotoxicology, and human health concerns.
	0	1	0	joe	Fipronil - overview	maybe later	2003	Fipronil: environmental fate, ecotoxicology, and human health concerns.
	1	0	0	joe	Fipronil - overview	well done	2008	The insecticide fipronil and its metabolite fipronil sulphone inhibit the rat
	0	0	1	mary	Fipronil - overview	good info	2008	The insecticide fipronil and its metabolite fipronil sulphone inhibit the rat

The Copy to Main button does just that. It copies all the Notes back to the Main sheet (overwriting what is there). Any colorization is retained.

The Import from RIS button lets the user choose a RIS file and import the fields into the corresponding columns on the Notes sheet.

The OA pdf link lets the user select rows and then the Sifter iterates through the rows and finds the open access pdf link for each, if they exist in Unpaywall. The link is placed in column Y.

The Highlight articles with curated chemicals button and the Notes curation sheets buttons will be discussed in other sections of this user guide.

The next section of More stuff form is for exporting records in various formats.

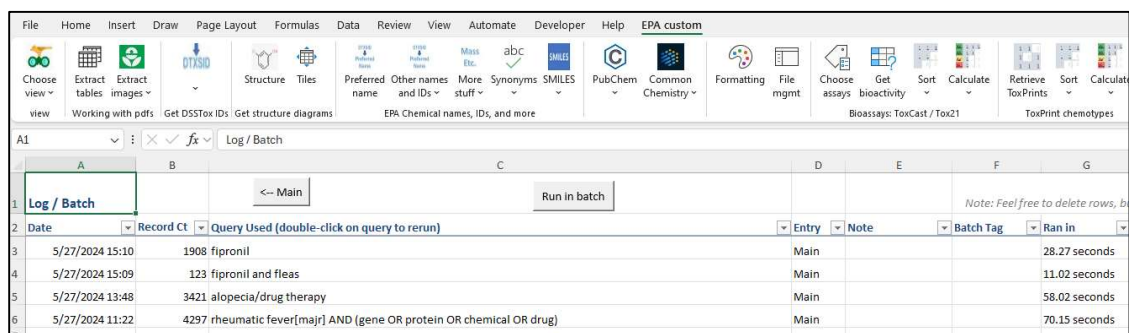
The Abstract Sifter allows the user to export articles from the Notes sheet to outside applications. On the More stuff button form there is a button labeled Export. (Figure E-10) Use the radio buttons at the bottom to select the format of the output. Next the user can choose to export all entries / rows on the Notes sheet or just selected rows. If the end-user selects PMIDs to be exported, the PMIDs will be formatted in the box. In this case, clicking on Next Step will copy the formatted PMIDs to the clipboard, ready to be pasted to the next application. In the case of PubMed, they should be pasted into the query box in PubMed Entrez. The options also include a RIS option for exporting citations to import into a citation management system like EndNote.

F. Log / Batch sheet

The Log sheet keeps track of the queries you have run. The Abstract Sifter routines insert a row into the sheet every time you complete a query. These queries can be viewed and rerun. To rerun a query, simply double-click on it. (Figure F-1)

The Batch Run option allows the user to run multiple queries and append the results from each on the Main sheet. To run in batch, select rows and click on the Run in Batch button. A Batch Tag can be added or modified on the Log sheet. This tag will be added to the Main sheet results and used to help summarize the results of multiple runs. This summary functionality is accessible on the Main sheet through the More things button. In section K of this user guide there is an in-depth discussion of the batch run capability with examples.

Delete any or all rows after Row 2 if you want to clear old entries.



	A	B	C	D	E	F	G
1	Log / Batch		<- Main	Run in batch			Note: Feel free to delete rows, b
2	Date	Record Ct	Query Used (double-click on query to rerun)	Entry	Note	Batch Tag	Ran in
3	5/27/2024 15:10	1908	fipronil	Main			28.27 seconds
4	5/27/2024 15:09	123	fipronil and fleas	Main			11.02 seconds
5	5/27/2024 13:48	3421	alopecia/drug therapy	Main			58.02 seconds
6	5/27/2024 11:22	4297	rheumatic fever[majr] AND (gene OR protein OR chemical OR drug)	Main			70.15 seconds

Figure F-1. View of the Log sheet.

G. Landscape sheet

The Landscape sheet provides an overview of the number of citations in PubMed to the user for a set of entities, for example, a list of chemicals or genes. Figure G-1 shows an example of a Landscape sheet built by a researcher interested in the toxicity of a particular set of chemicals. Let's take a look at that first. Queries designed to find the chemicals of interest are entered into Column C and in this case, a short version of the chemical name is in Column B. The queries in Row 3 are typical ones used in searching for articles about different kinds of chemical toxicity. We will refer to these queries as subject matter queries. (Note: Column A on the Landscape sheet is often hidden. Go ahead and unhide it and use it when you have a DSSTox chemical identifier.)

The premise behind the design of the Landscape sheet is very simple: PubMed queries will be built by taking the values in Column C (in this example chemical names and corresponding CAS numbers) and appending this query text to the subject matter query text in Row 3 with an " AND " in between the two query parts.

B		C		E	F	G	H	I	J
1	Landscape View								
2		Update Article Counts	View / hide queries	Heat Map by column	Heat Map by row				
3									
4									
5	Preferred Name	Chemical / Entity query		Genetox	Cancer	ReproTox	NeuroTox	DevTox	Respiratory sensitization
6	PERC/TCE	Tetrachloroethylene[majr]		42	1152	39	68	14	22
7	Tripropylene glycol	Tripropylene glycol OR 24800-44-0		4	7	0	0	0	0
8	Tetrachlorophthalic anhydride	Tetrachlorophthalic anhydride		1	0	0	0	0	11
9	Linalool	Linalool		41	78	0	0	0	20
10	TBBPA	Tetrabromobisphenol A OR TBBPA		16	60	0	0	0	3
11	Dronabinol / THC	Dronabinol		61	470	154	2034	55	130
12	TPHP	Triphenyl phosphate OR triphenylphosphate		9	22	27	28	32	4
13	BDE-100 / PBDE	pentabrominated diphenyl ether 100 OR BDE-100 OR 5436-43-1 OR 2,2',4,4'-Tetr		97	240	350	351	234	13
14	Styrene	Styrene		1041	1942	102	533	79	247
15	PCB126	3,4,5,3',4'-pentachlorobiphenyl		22	80	58	53	73	6
16									

Figure G-1. Structure of Landscape sheet

To illustrate, we will double-click on the cell with the arrow pointer in Figure G-2. When we double-click on this cell this tells the Abstract Sifter to take the query text in Column C about Linalool and append it to query text designed to find citations about reproductive toxicity. Figure G-3 shows the constructed query. We can then click on Submit and the query gets sent to PubMed and we can then see the results on the Main sheet. The number of articles retrieved from PubMed is 5. That count is placed in the corresponding Landscape cell that we just clicked on.

	chromosome aberrations OR	carcinogen* OR precancerous	abnormal OR adverse effects))	chemically induced OR
Subject queries:				
Chemical / Entity query	Genetox	Cancer	ReproTox	NeuroTox
Tetrachloroethylene[majr]	42	1152	39	6
Tripropylene glycol OR 24800-44-0	4	7	0	0
Tetrachlorophthalic anhydride	1	0	0	0
Linalool	41	78	5	1
Tetrabromobisphenol A OR TBBPA	16	60	48	1
Dronabinol	61	470	134	263
Triphenyl phosphate OR triphenylphosphate	9	22	27	3

Figure G-2. Double-click on article count cells to retrieve citations.

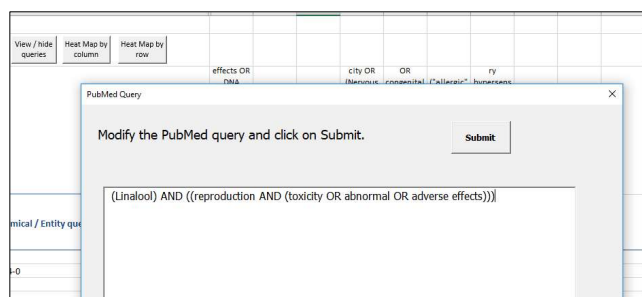


Figure G-3. A query is constructed with values in column C appended with AND to values in row 3.

Now let's add to the Landscape sheet. Figure G-4 shows how we added a new chemical to the list: aspirin. To find out the article counts for aspirin, select empty cells on the same row as aspirin, then click on Update Article Counts button. Excel will build each query from the aspirin part and the subject matter part and send each query to PubMed to find out how many citations satisfy the query. (Figure G-5) The article counts are placed in the corresponding cells. To run the query and retrieve the results, just double-click on any of the article count cells.

article counts either by column or by row. Try then out for a quick way to improve visualization. (Figure G-7)

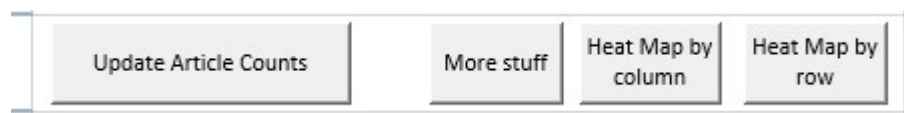


Figure G-7. Buttons on the Landscape sheet include formatting actions.

The More stuff button opens a form that contains currently two options. The Show/Hide queries just hides the query row for you in case you don't know how.

The second button will take in turn each of the selected citation cells, build a query for each cell and send those queries to the Log sheet. (See Figure G-8) Here the queries can be run and rerun individually by double-clicking on a row or they can be run in batch by selecting rows and clicking on the Run in batch button. This action will run each query and append the results on the Main sheet. For more detail on this feature, see section K2 in this user guide where you'll find lots of detail.

These 3 cells are selected then Send to Log clicked. Look at what goes to the Log sheet.

Next, select the 3 rows and then click on Run in batch. The 3 queries will be run and the results appended to each other on the Main sheet.

Figure G-8. Select cells on the Landscape sheet and Send to Log via the More stuff button. The queries can be run on the Log sheet individually or in batch.

H. Sample_Queries, Pathway_Queries, CuratedLists Sheets

The Sample_queries and pathway_queries sheets function in a very similar way. We will use the Sample_queries sheet as an example in this user guide. Both sheets contain a number of sample subject matter queries that the end user can use as a starting point for building a Landscape view of a set of entities. It is important to stress that these queries should be considered starting points. Feel free to add to them and modify what is there. If you do change a query, check it by either copying and pasting into PubMed or double-clicking on it to have it sent to PubMed. PubMed has some error checking capabilities that might catch missing or misplaced parentheses and common errors like that.

Let's see how to put these queries to use. First, we will clean off the old subject matter queries by deleting columns E-L on the Landscape sheet. (You can let the previous work stay if you wish.) Next, on the Sample_Queries sheet we will select rows with queries of interest then we click on the button Send Queries to Landscape (Figure H-1).


Sample Queries		Note: these are starting points ... please expand and customize	Send queries to Landscape 
Category	Heading	Query (double-click to see how the query looks to PubMed)	
Methods	In vitro	In Vitro Techniques[mh] OR cell culture or "in vitro"	
Mixtures	Mixtures	(Drug synergism[mh] OR cocarcinogenesis OR pesticide synergists[mh] OR mixture[tiab] OR mixtures[tiab] OR Drug Antagonism[n	
Medicine	Clinical trials	((clinical[Title/Abstract] AND trial[Title/Abstract]) OR clinical trial[Publication Type])	
Medicine	Clinical trials in children	((children OR child OR infants) AND human) AND ((clinical[Title/Abstract] AND trial[Title/Abstract]) OR clinical trial[Publication Typ	
Medicine	Obesity	(obesity OR obese OR adipose OR overweight OR adipogenesis OR adipose tissue)	
Toxicity	Genetox	(dna/drug effects OR DNA Damage OR chromosome aberrations OR genotoxicity OR micronucleus OR DNA Repair OR mutagenicity	
Toxicity	Cancer	neoplasms or cancer OR carcinogen* OR precancerous	
Toxicity	ReproTox	(reproduction AND (toxicity OR abnormal OR adverse effects))	
Toxicity	NeuroTox	(neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR brain OR behavior) AND drug effects)	
Toxicity	DevTox	((toxicity OR congenital abnormalities OR Prenatal Exposure Delayed Effects) AND (fetus OR embryo OR embryonic development C	
Toxicity	Skin sensitization	("allergic" AND "contact" AND dermatitis) OR Dermatitis, Allergic Contact[mh]	
Toxicity	Respiratory sensitisation	(Respiratory hypersensitivity OR respiratory sensitization OR Bronchial Hyperreactivity OR Respiration Disorders OR Respiratory Tr	
Toxicity	DNT	((Brain OR central nervous system OR "CNS" OR "neural tube" OR spinal cord OR spina bifida OR Nervous System Diseases or Neura	
Use	Pharmaceutical	"therapeutic use" OR "therapeutic use"[subheading] OR pharmacologic actions[mh] OR drug therapy	
Use	Pesticide	pesticide OR insecticide OR rodenticide OR fungicide	
Use	Cosmetics	cosmetics OR beauty	
Use	Explosive Agents	Explosive Agents OR explosive OR explosives	
Use	Food	food OR diet OR beverage OR nutrition OR eating	
Use	Surface-acting	Antifoaming OR Anti-foaming OR detergent OR detergents OR soap OR detergent OR surfactant	
Use	Dye/coloring	dye OR "coloring agent" OR pigment OR pigments	
Use	Fertilizer	fertilizer OR fertilize	
Use	Solvents	solvents OR solvent	

Figure H-1. Selecting rows with queries of interest.

Our Landscape sheet then looks like Figure H-2. The query terms have been written to Row 3 and the query heading written to Row 5.

A	B	C	E	F	G	H	I	J	K	L
Abstract Sifter	Landscape View									
v5.6		Update Article Counts	More stuff	Heat Map by column	Heat Map by row					
			diet OR beverage OR nutrition OR	ing OR Anti-foaming OR detergent	(dna/drug effects OR DNA Damage OR	ms or cancer OR carcinogen* OR	ction AND (toxicity OR abnormal	icity OR (Nervous system diseases and	OR congenital abnormalities OR	"AND "contact" And dermatitis) OR
		Subject queries:								
(optional) DSSTOX link to Dashboard	Preferred Name	Chemical / Entity query	Food	Surface-actin	Genetox	Cancer	ReproTox	NeuroTox	DevTox	Skin sensitization
DTXSID1021322	Disulfiram	97-77-8 OR Disulfiram								
DTXSID6024337	Thiobencarb	28249-77-6 OR Thiobencarb OR benthioicarb								
DTXSID3023556	Retinol	68-26-8 OR Retinol OR Vitamin A								
DTXSID2022880	Danazol	17230-88-5 OR Danazol								
DTXSID9020453	Dieldrin	60-57-1 OR Dieldrin								
DTXSID7032638	Pyraclostrobin	175013-18-0 OR Pyraclostrobin OR pyrachlostrobin								
DTXSID8024151	Imazalil	35554-44-0 OR Imazalil OR enilconazole								
	Linalool	Linalool								
	Styrene	Styrene[majr]								
	TPHP	Triphenylphosphate OR "triphenyl phosphate"								
	PERC/TCE	Trichloroethylene[majr]								
	Dronabinol/THC	Dronabinol								

Figure H-2. New queries on Landscape sheet.

Next, we select the article count area and then click on *Update Article Counts*. (Figure H-3)

Abstract Sifter	Landscape View									
v5.6		Update Article Counts	More stuff	Heat Map by column	Heat Map by row					
		Subject queries:								
(optional) DSSTOX link to Dashboard	Preferred Name	Chemical / Entity query	Food	Surface-actin	Genetox	Cancer	ReproTox	NeuroTox	DevTox	Skin sensitization
DTXSID1021322	Disulfiram	97-77-8 OR Disulfiram								
DTXSID6024337	Thiobencarb	28249-77-6 OR Thiobencarb OR benthioicarb								
DTXSID3023556	Retinol	68-26-8 OR Retinol OR Vitamin A								
DTXSID2022880	Danazol	17230-88-5 OR Danazol								
DTXSID9020453	Dieldrin	60-57-1 OR Dieldrin								
DTXSID7032638	Pyraclostrobin	175013-18-0 OR Pyraclostrobin OR pyrachlostrobin								
DTXSID8024151	Imazalil	35554-44-0 OR Imazalil OR enilconazole								
	Linalool	Linalool								
	Styrene	Styrene[majr]								
	TPHP	Triphenylphosphate OR "triphenyl phosphate"								
	PERC/TCE	Trichloroethylene[majr]								
	Dronabinol/THC	Dronabinol								

Figure H-3. Selecting the cells for article counts.

Once the article counts are populated, we click on Heat Map by Row and then on Hide queries. Our resulting Landscape view looks like Figure H-4. To run the query and retrieve the results, just double-click on any of the article count cells.

Note that the Pathway_queries sheet works the same way: select the rows of interest and click on the button Send queries to Landscape. To add your own queries to either the pathway or sample queries sheets, just enter in new rows. Use the existing rows as a template to indicate where headers go and where the query text goes. Be mindful of parentheses and be sure to test the queries.

I. TermMap Sheet

Term mapping helps rank a corpus of articles on the Main sheet so that those articles of interest and relevance rise to the top.

For instance (and sticking with our familiar fipronil example) a scientist researching fipronil wants to find articles about rodent toxicity studies. Term Mapping can be used to find that type of article quickly and easily. If then the next day the researcher wants to read about environmental fate of fipronil, Term Mapping can be used to rank the corpus putting articles about environmental fate at the top.

Let's pretend we are this scientist and walk through how to do it. First, we'll rerun the fipronil query on the Main sheet. Then we'll go to the Term Map sheet. In this example, the sheet has been cleared, but don't worry if it hasn't.

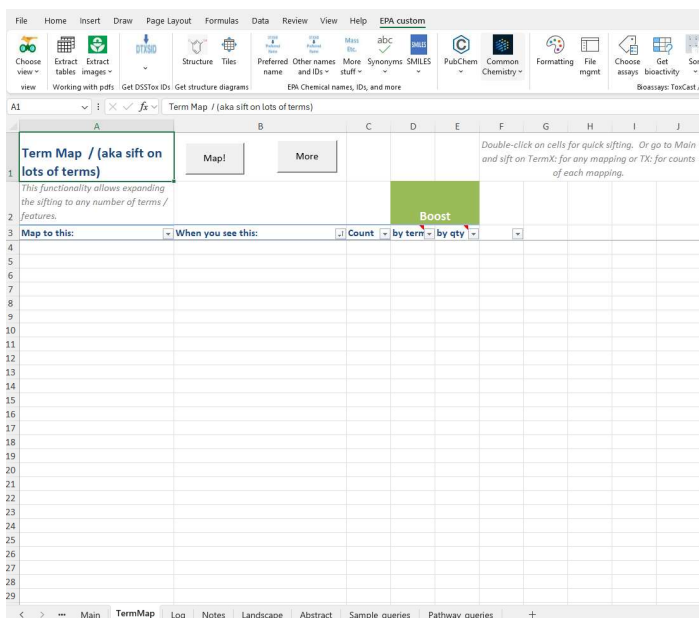


Figure I-1. TermMap sheet when it is empty.

A	B	C
Term Map / (aka sift on lots of terms)	Map! More	
This functionality allows expanding the sifting to any number of terms / features.		
Map to this:	When you see this:	Count
rodent	mice	57
rodent	rats	107

Figure I-2. Sample data on TermMapping sheet.

This sheet is where the terms of interest (in data mining these are called features) are entered. They go in column B with a mapped to term (just a short higher level term) in column A. Let's type in *mice* and *rats* in column B and *rodent* in column A.

A	B
Term Map / (aka sift on lots of terms)	Map! More
This functionality allows expanding the sifting to any number of terms / features.	
Map to this:	When you see this:
rodent	mice
rodent	rats

Then click on the *Map!* Button. Clicking on the *Map!* Button directs the Abstract Sifter to read through the titles, abstracts, and keywords of the citations on the Main sheet and look for the string of characters “mice”

and “rats”. The number of articles that have each term will be counted and the counts put into column C.

In addition to the counts, the terms have been appended to the value in column L on the Main sheet. This is the column – normally hidden - where the title, abstract, and keywords are stored.

Here's how this cell is modified: if any of the column B terms was found in the title and abstract, then “TERMX:” was appended to the abstract. For each column B term found, “TX: “ and the column A value plus the Column B value were stuck on the end of the abstract.

To see an example that will make this clearer ... see Figure xx of an abstract on the Abstract Sheet. At the end you see “TERMX: |TX:Rodent:mice |TX:Rodent:rats. That's what the term mapping feature added. Caution when using terms like “rat” and “mice” as those strings of characters can be parts of larger words and will introduce error in the counts.

Abstract with highlights	
PMID:	33993955
DocID:	33993955
Title:	The effects of fipronil on emotional and cognitive behaviors in mammals.
Title and Abstract:	<p>The effects of fipronil on emotional and cognitive behaviors in mammals. ABSTRACT: Fipronil is a phenylpyrazole insecticide that is widely used as a pesticide and a veterinary drug, although studies suggest that it could be toxic to mammals. The objectives of this study were to examine the pharmacokinetic profile of fipronil in mice, dogs, and cats, and to evaluate its effects on emotional and cognitive behaviors of dogs and cats using the data obtained from mice. The assessment of in vivo kinetics of fipronil was conducted in mice and dogs. We also performed behavioral tests (elevated plus-maze and Y-maze) and measured the levels of neurotransmitters in mice exposed to fipronil. In addition, the in vitro metabolism of fipronil were evaluated using liver microsomes of rats, mice, dogs, and cats. The results revealed that fipronil is distributed throughout the body (blood, brain, adipose tissue, and liver) of mice after dermal application. It was metabolized to fipronil sulfone primarily in the liver. The data on kinetics show that both fipronil and fipronil sulfone have a longer half-life in dogs and cats than in mice. The behavioral tests indicated that fipronil and fipronil sulfone could affect emotional and cognitive behaviors and alter the levels of neurotransmitters (dopamine in the striatum and serotonin in the hippocampus) in mice. Furthermore, we found that dogs and cats have a low ability to metabolize fipronil than mice and rats. However, further comprehensive studies are needed to determine whether fipronil affects the emotional and cognitive behaviors when administered to dogs and cats. To the best of our knowledge, this is the first study to examine the pharmacokinetic data and verify the effects of fipronil on emotional and cognitive behaviors of dogs and cats using the data obtained from mice. Copyright © 2021 Elsevier Inc. All rights reserved. KEYWORDS: MeSH Chemical: Insecticides Pyrazoles fipronil MeSH: Animals Cat Diseases/majr Cats Cognition Dog Diseases/majr Dogs Insecticides/majr/toxicity Mammals Mice Pyrazoles Rats Other:Fipronil Fipronil sulfone Mammal behavior Mammal cognition Veterinary drugs TERMX: TX:rodent:mice TX:rodent:rats TXTot:rodent</p>

Figure I-3. How term mapping appends values to the abstract stored in Excel.

In the Abstract Sifter, if text exists in the abstract column that means it can be sifted on. There are three new ways to sift if you map terms.

1. On the Main sheet, in cell B3, C3, or D3 - Sift on **TERMX**: ... this will show you any citation that has a least one of the column B terms.
2. On the Main sheet, cell B3, C3, or D3 - Sift on **TX**: ... this will count the number of terms from column B found in the citation.
3. Sift on a value in column A or column B. To do this quickly, stay on the TermMap sheet and double-click on a term in either A or B and see what happens. (What does happen? The Sifter copies the term to the Main sheet cell B3, lets the Main sheet sift the results, and then sorts by B3. It happens fast, but it is very simple.)

Now let's look at extending the list of terms. Click on the **More** button on the Term Map sheet. This action presents a list of actions. The bottom part of the form has button that will cause the Abstract Sifter to populate the column A and B term area with sample lists that can be edited and enhanced.

Let's click on Rodent Studies and see what happens. (Figure I-4)

Figure I-4. Selecting Rodent Studies and getting stored suggestions for terms.

A long list of terms having some kind of connection to rodent studies appears. Consider it a starting point to be added to or deleted from as needed. Then click on **Map!** to get the counts.

Suppose you want to find the article on the Main sheet that is most about rodents. In our definition that means the one that has the most features from our rodent feature (term) lists. Double-click on the rodent on the TermMap sheet. You'll be taken to the Main sheet, the sifter cell filled in and rows sorted (Figure I-5). The records on Main are now sorted by the number of features they have connecting them to the Rodent feature/term set.

	A	B	C	D	E	F	G
	Abstract Sifter		Query PubMed		PubMed query run: fipronil		
1	Version 8.0		Your sifter terms and frequency counts				
2		TX:Rodent	mice	mg	Score		
3	PMID				Pub Y-	Title	
4	23962444	9	9	0	18	2013 The nuclear receptors pregnane X receptor and constitutive androsta	
5	27371222	9	6	5	20	2017 Fipronil-induced genotoxicity and DNA damage in vivo: Protective eff	
6	12442503	8	4	5	17	2003 Fipronil: environmental fate, ecotoxicology, and human health conce	
7	33201557	7	8	0	15	2021 Efficacy of a combined insecticide-rodenticide product on ectoparasit	
8	38254069	7	7	2	16	2024 Fipronil disturbs the antigen-specific immune responses and GABAergic	
9	29368479	7	5	3	15	2016 [Genotoxic effects of pesticide fipronil in somatic and generative cell	
10	35511311	7	3	16	26	2022 Sensitivity of the stripe-faced dunnart, <i>Sminthopsis macroura</i> (Gould)	
11	29854616	7	0	3	10	2018 Assessment of fipronil toxicity on wistar rats: A hepatotoxic perspect	
12	31428841	7	0	2	9	2019 Blood pharmacokinetic of 17 common pesticides in mixture following	
13	31873888	7	0	4	11	2020 Synergistic antioxidant effects of resveratrol and curcumin against fip	
14	34493330	6	15	0	21	2021 Efficacy of low-dose fipronil bait against blacklegged tick (<i>Ixodes sca</i>	
15	16836887	6	15	2	23	2006 [An experimental study on acute poisoning by fipronil in mice and its	
16	11812623	6	7	10	23	2002 Development of a mouse model to determine the systemic activity of	
17	17311043	6	3	0	9	2007 Agitation, pruritus, and ectoparasites on a rat and a mouse.	
18	24640131	6	3	3	12	2013 [The systemic effects of imidacloprid and fipronil on <i>Xenopsylla cheo</i>	
19	27166213	6	0	2	8	2017 Perinatal exposure to insecticide fipronil: effects on the reproductive	
20	31881178	6	0	2	8	2020 Transcriptomic modifications of the thyroid gland upon exposure to p	

You may want to add weighting to the terms. The Boost columns let you do this. Use this feature sparingly. Here's how. Let's say you have a term feature that is really important and will help discriminate between citations that are really of interest and those that are of less interest. Let's say we want to boost a species of mouse. It can be done by term. Let's see what happens when you put a 1 in the by term column (Figure I-6) then click on *Map!*.

A										
Abstract with highlights		← Main		→ (shift and left arrow)		→ (shift and right arrow)		Add Note	See existing notes	Like this?
PMID:	27371222									
DocID:	27371222									
Title:	Fipronil-induced genotoxicity and DNA damage in vivo: Protective effect of vitamin E.									
Title and Abstract:	<p>Fipronil-induced genotoxicity and DNA damage in vivo: Protective effect of vitamin E. ABSTRACT: Fipronil, an insecticide of the phenylpyrazole class has been classified as a carcinogen by United States Environmental Protection Agency, yet very limited information is available about its genotoxic effects. Adult male and female animals were gavaged with various doses of fipronil (2.5, 12.5, and 25 mg/kg body weight [bw]) to evaluate micronucleus test (mice), chromosome aberration (CA), and comet assay (rats), respectively. Cyclophosphamide (40 mg/kg bw; intraperitoneal) was used as positive control. Another group of animals were pretreated with vitamin E orally (400 mg/kg bw) for 5 days prior to administration of fipronil (12.5 mg/kg). Fipronil exposure in both male and female mice caused significant increase in the frequency of micronuclei (MN) in polychromatic erythrocytes. Similarly, structural CAs in bone marrow cells and DNA damage in the lymphocytes was found to be significantly higher in the male and female rats exposed to fipronil as compared to their respective controls. The average degree of protection (male and female animals combined together) shown by pretreatment of vitamin E against fipronil-induced genotoxicity was 63.28% CAs; 47.91% MN formation; and 74.70% DNA damage. Findings of this study demonstrate genotoxic nature of fipronil regardless of gender effect and documents protective role of vitamin E. KEYWORDS: MeSH Chemical: Antiparasitic Agents Insecticides Protective Agents Pyrazoles Vitamins Vitamin E Fipronil MeSH: Animals Antiparasitic Agents/major toxicity Bone Marrow Cells/drug effects Comet Assay DNA Damage Female Insecticides/major toxicity Lymphocytes/drug effects Male Mice Micronuclei, Chromosome-Defective/chemically induced Micronucleus Tests Protective Agents/major pharmacology Pyrazoles/major toxicity Rats Rats, Sprague-Dawley Vitamin E/major pharmacology Vitamins/major pharmacology Other chromosome aberrations comet assay fipronil micronucleus test rats and mice vitamin E TERMX: TX:Rodent:control TX:Rodent:dose TX:Rodent:gavage TX:Rodent:mg TX:Rodent:mice TX:Rodent:mice TX:Rodent:rats TX:Rodent:rats TX:Rodent:Sprague-Dawley TX:Rodent</p>									

Figure I-5. Term mapping gives a new way to rank by relevancy.

Map to this:		When you see this:		Art Ct	by term	by qty
fate		Dehalobacter		0		
fate		Escherichia		73		
fate		Acinetobacter		7	1	
fate		catalytic oxid		0		

Figure I-6. Boosting the score by term.

When the mapping is done, double-click on *Acinetobacter* to be sent to the Main sheet and look at one of abstracts by double-clicking. *Acinetobacter* has boosted the score of the citation, causing it to be ranked higher.

PMID	Acinetobacter	fate	transport	Score	Pub Yr	Title
27483290	6	3		9	2016	5-Episinuleptolide Decreases the Expression of the Extracellular Matrix in Early Biofilm Formation of Multi-Drug Resistant <i>Acinetobacter baumannii</i>
31923557	3	5		8		Antimicrobial profiling of coral reef and sponge associated bacteria from southeast coast of India.
23777289	3	4		7		Coral-associated bacteria, quorum sensing disruptors, and the regulation of biofouling.
23205044	3	4		7		Ramarinolides A-D: antimicrobial butenolides isolated from the mushroom <i>Ramaria ruficincta</i>

gents/majr/chemistry/isolation & purification/pharmacology | Biofilms/majr/drug effects | Biological Products/Control | Diterpenes/majr/chemistry/isolation & purification/pharmacology | Drug Resistance, Multiple, Bacteria | Effects/ultrastructure | Genes, Bacterial/drug effects | Humans | Levofloxacin/pharmacology | Microscopy, Electron | *Acinetobacter baumannii* | TERMX: TX:fate:Acinetobacter | TX:fate:Acinetobacter | TX:fate:persist

Figure I-7. Illustration of boost by term effects on the Mainsheet and abstract sheet. An additional mapping gets added to the abstract causing the term score to increase ranking.

To see what happens when the *by qty* option is chosen, put a 1 next to Acinetobacter, then click on Map!. Then double-click on Acinetobacter and pick a citation to look at. Here's an example. The term Acinetobacter is appended to the abstract for each occurrence. This really boosts the term score for the citation.

This functionality allows expanding the sifting to any number of terms / features.				Boost	
Map to this:	When you see this:	Art Ct	by term	by qty	
9 fate	Dehalobacter	0			
10 fate	Escherichia	73			
11 fate	Acinetobacter	7		1	
12 fate	antibiotic resistance	0			

Title: 5-Episinuleptolide Decreases the Expression of the Extracellular Matrix in Early Biofilm Formation of Multi-Drug Resistant *Acinetobacter baumannii*.

Title and Abstract: 5-Episinuleptolide Decreases the Expression of the Extracellular Matrix in Early Biofilm Formation of Multi-Drug Resistant *Acinetobacter baumannii*. ABSTRACT: Nosocomial infections and increasing multi-drug resistance caused by *Acinetobacter baumannii* have been recognized as emerging problems worldwide. Moreover, *A. baumannii* is able to colonize various abiotic materials and medical devices, making it difficult to eradicate and leading to ventilator-associated pneumonia, and bacteremia. Development of novel molecules that inhibit bacterial biofilm formation may be an alternative prophylactic option for the treatment of biofilm-associated *A. baumannii* infections. Marine environments, which are unlike their terrestrial counterparts, harbor an abundant biodiversity of marine organisms that produce novel bioactive natural products with pharmaceutical potential. In this study, we identified 5-episinuleptolide, which was isolated from *Sinularia leptoclados*, as an inhibitor of biofilm formation in ATCC 19606 and three multi-drug resistant *A. baumannii* strains. In addition, the anti-biofilm activities of 5-episinuleptolide were observed for Gram-negative bacteria but not for Gram-positive bacteria, indicating that the inhibition mechanism of 5-episinuleptolide is effective against only Gram-negative bacteria. The mechanism of biofilm inhibition was demonstrated to correlate to decreased gene expression from the *pgaABCD* locus, which encodes the extracellular polysaccharide poly-β-(1,6)-N-acetylglucosamine (PNAG). Scanning electron microscopy (SEM) indicated that extracellular matrix of the biofilm was dramatically decreased by treatment with 5-episinuleptolide. Our study showed potentially synergistic activity of combination therapy with 5-episinuleptolide and levofloxacin against biofilm formation and biofilm cells. These data indicate that inhibition of biofilm formation via 5-episinuleptolide may represent another prophylactic option for solving the persistent problem of biofilm associated *A. baumannii* infections. KEYWORDS: MeSH Chemicals: 5-episinuleptolide [Antibacterial Agents] [Biological Products] [Diterpenes] [beta-Glucans] [poly-N-acetyl-1,6-glucosamine] [Levofloxacin MeSH: *Acinetobacter* Infections/majr/microbiology/prevention & control] | *Acinetobacter baumannii*/majr/drug effects/physiology/ultrastructure | Animals | Antibiotics/majr/chemistry | Anti-Bacterial Agents/majr/chemistry/isolation & purification/pharmacology | Biofilms/majr/drug effects | Biological Products/isolation & purification/pharmacology | Cross Infection/majr/microbiology/prevention & control | Diterpenes/majr/chemistry/isolation & purification/pharmacology | Drug Resistance, Multiple, Bacteria/majr/drug effects | Equipment Contamination | Extracellular Matrix/drug effects/ultrastructure | *Escherichia coli*/majr/drug effects | Humans | Levofloxacin/pharmacology | Microscopy, Electron, Scanning | beta-Glucans/meSH Chemicals: 5-episinuleptolide | biofilm | multi-drug resistant *A. baumannii* | TERMX: TX:fate:Acinetobacter | TX:fate:Acinetobacter | TX:fate:Acinetobacter | TX:fate:Acinetobacter | TX:fate:persist

Figure I-8. Illustration of boost by quantity. An additional mapping gets added to the abstract for each occurrence of the mapping features causing the term score to increase ranking.

The EPA has assembled a list of common term feature lists. Some of

Things you can do

Map! <- Click here to map your terms to the citations on the Main sheet with totals

Wondering how you might use term mapping? Here are some examples. Try them out ...

- Rodent studies <- Click here to populate an example of term expansion for rodent studies. Use them on a query for chemicals such as fipronil or triphenyl phosphate.
- Coral ontology <- Click here to populate an example of term mapping with a play ontology for coral. Use it on a query about coral such as "coral AND (pollution OR mortality OR toxicity)"
- Heart development gene ontology <- Click here to populate an example of term expansion with a sample Mouse Genome Informatics gene ontology for heart development. Use it on a query like "embryo AND heart development AND gene expression"
- Environmental Fate <- Click here to populate the mapping list with term features associated with environmental fate
- Phys chem <- Click here to populate the mapping list with term features associated with physicochemistry
- More feature lists <- Click here to see more list options

A drop-down list of term feature sets is available by clicking on **More feature lists** button.

Select Feature List

Select a list and click on Load this ...

- Mature terms
- Chemical identification
- Emergencies
- Microglia in development
- Water
- Media
- Epiblast genes
- PPAS

Load this

Figure I-9. How to get stored starting point feature lists.

them are available through the buttons shown on the More button. However, a growing list is available by clicking on *More feature lists* button. This action brings up a drop-down menu of feature lists constructed here at the EPA. (Figure I-9) Try them out, and if there is a set of term features you would like to contribute to the community, let us know.

J. MeSHMine sheet

MeSH mining is a feature to help answer the needs of researchers who ask questions like these:

- I ran a query on Histamine H2 antagonism. Can I get a list of all the chemicals annotated in those PubMed records?
- I ran a query on unfolded protein response and chemical toxicity. I'd like to find some chemicals that cause this response. Is there a fast way to do this?
- I ran a query on retinal disease. Can I get a list of genes and proteins annotated in my resulting article set on Main?
- I ran a query on biosolids. I want to see a list of chemicals found in biosolids.
- I ran a query on fipronil. I want to know the genes and proteins associated with this chemical.

As the above list shows, step one in using this new feature is to run a query to the Main sheet. Let's walk through an example.

Let's say we ran the query *(Anthozoa OR coral) AND (chemical toxicity OR drug effects)*.

Running this query brings back 2,801 records to the Main sheet. Next, click on the button More things on the top of the Main sheet. Click on Extract MeSH Substances. The next form asks whether you want to extract all annotated chemicals or just the major topics. If you don't know what this means, read about annotation at the MeSH web site: <https://www.nlm.nih.gov/mesh/meshhome.html> or https://www.nlm.nih.gov/mesh/intro_retrieval.html

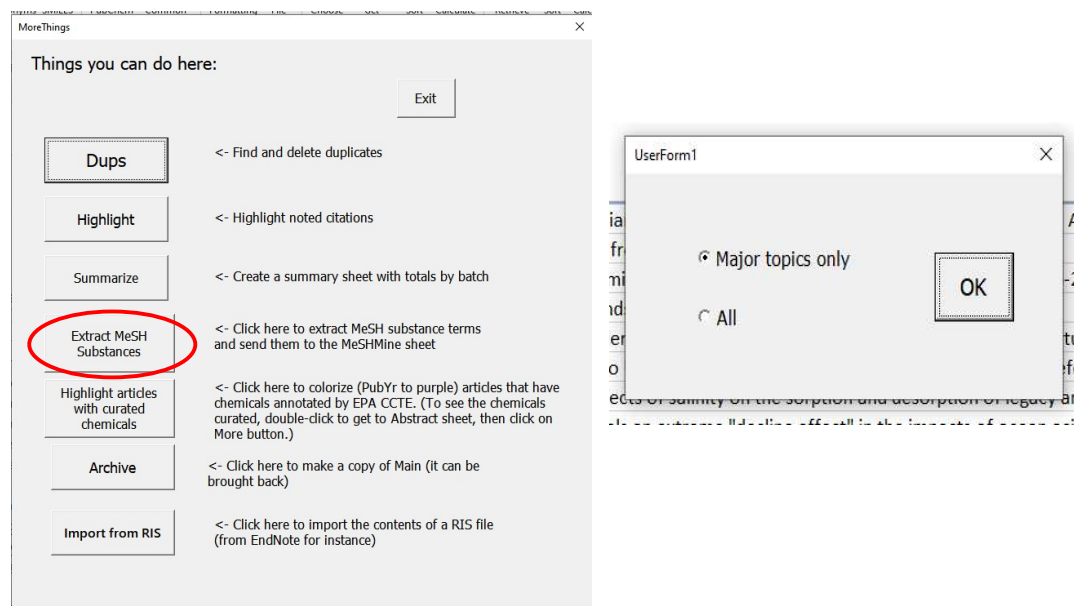


Figure J-1. Buttons for extracting MeSH substances. Try both Major topics and all and observe the differences.

In a few minutes or seconds (depending on the size of the collection on Main and your internet speed) you will be taken to the MeSHMine sheet and a total displayed.

Let's see what's on the MeSHMine sheet (Figure I-45).

A		C	D	E	F	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
MeSHMine		PubMed query run: (Anthozoa OR coral) AND (chemical toxicity OR drug effects).																					
Chemicals, proteins, and more: double-click on row to sift on substance name		D01 D02 D03 D04 D05 D06 D08 D09 D10 D12 D13 D20 D23 D25 D26 D27 Oth																					
<div>More</div>		Art Ct	Tox / AE	Ther Use	Family		Inorganic	Organic	Heterocyclic	Polycyclic	Macro	Hormone, etc.	Enzymes	Carbs	Lipids	Prots, genes, AA	Nucleic acids	Mixtures	Biological factors	*Med/dental materials	*Pharm preparations	Actions/uses	Other
MeSH Substance (hyperlinked)						Mapped to Name																	
4	Calcium	17	0	1	0	Calcium	1												1				
5	Calcium Channels, L-Type	2	0	0	1	Calcium Channels, L-Type		1								1							
6	Microplastics	9	7	0	1	Microplastics					1									1			
7	Polystyrenes	1	1	0	1	Polystyrenes		1			1									1			
8	Anti-Bacterial Agents	100	7	2	1	Anti-Bacterial Agents																1	
9	Ampicillin	1	0	0	0	Ampicillin		1	1														
10	Petroleum	24	20	0	0	Petroleum											1						
11	Water Pollutants, Chemical	274	208	1	1	Water Pollutants, Chemical																	1
12	corexit 9500	1	1	0	0	Lipids									1								
13	Lipids	5	1	0	1	Lipids									1								
14	Surface-Active Agents	9	6	0	1	Surface-Active Agents																	1
15	Antineoplastic Agents	198	8	22	1	Antineoplastic Agents																	1
16	Prodrugs	1	0	0	1	Prodrugs																1	
17	Iron	2	1	0	0	Iron	1																
18	Oxygen	9	1	0	0	Oxygen	1																
19	Diterpenes	306	14	10	1	Diterpenes		1															
20	Antivenins	11	0	5	1	Antivenins										1		1					
21	Elapid Venoms	63	38	0	1	Elapid Venoms												1	1				
22	micurus venom	28	13	0	0	Elapid Venoms												1	1				

Figure J-2. MeSHMine sheet.

Column A has the substance term. Try sorting on it. It can be a MeSH heading or MeSH supplemental concept (again, go to the PubMed help to find out the difference.) Each name is hyperlinked to the MeSH browser. You can click on the name to go to the browser to find out more.

The ArtCt column has the total number of articles on the Main sheet in which the MeSH terms was annotated. (Note – this is not the same as occurrences in the text of the abstract.) The Tox/AE column counts the number of articles on Main in which the substance appeared with the co-annotation toxicity or adverse effects or poisoning. The next column titled Ther Use counts the articles in which the substance is co-annotated with the subheading therapeutic use. Try sorting on these columns.

The column entitled Family contains a flag: 1 = family name ; 0 = not a family name. If you are interested in individual chemicals only, you can sort by this column and delete the rows with a 0 in the Family column. Or you can click on the *More* button above, then the *Delete family terms* button and remove them.

The rest of the columns contain some more information about the chemical and its place in the MeSH tree structure. If you are not familiar with the MeSH tree structure, go here and browse: <https://meshb.nlm.nih.gov/treeView> These columns can be used to sort the list. For instance, if you are only interested in genes, proteins, peptides, amino acids, then sort (descending) on column S.

	A	C	D	E	F	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
1	MeSHMine																							
2	Chemicals, proteins, and more: double-click on row to sift on substance name																							
		Art Ct	Tox / AE	Ther Use	Family																			
3	MeSH Substance (hyperlinked)					Mapped to Name																		
4	Calcium	17	0	1	Calcium																			
5	Calcium Channels, L-Type	2	0	0	Calcium Channels, L-Type																			
6	Microplastics	9	7	0	Microplastics																			
7	Polystyrenes	1	1	0	Polystyrenes																			
8	Anti-Bacterial Agents	100	7	2	Anti-Bacterial Agents																			
9	Ampicillin	1	0	0	Ampicillin																			
10	Petroleum	24	20	0	Petroleum																			
11	Water Pollutants, Chemical	274	208	1	Water Pollutants, Chemical																			
12	corexit 9500	1	1	0	Lipids																			
13	Lipids	5	1	0	Lipids																			
14	Surface-Active Agents	9	6	1	Surface-Active Agents																			
15	Antineoplastic Agents	198	8	22	Antineoplastic Agents																			
16	Prodrugs	1	0	0	Prodrugs																			
17	Iron	2	1	0	Iron																			
18	Oxygen	9	1	0	Oxygen																			
19	Diterpenes	306	14	10	Diterpenes																			
20	Antivenins	11	0	5	Antivenins																			
21	Elapid Venoms	63	38	0	Elapid Venoms																			
22	micurus venom	28	13	0	Elapid Venoms																			

Figure J-3. Filtering the MeSHMine sheet by Family = 0 and Organic = 1. Note that several columns used by the Abstract Sifter code are hidden here. Column I can be hidden as well, if desired.

Figure I-46 shows the results when the Family flag = 0 (not a family term) and the Organic flag is set to 1. (Note – there are a number of columns on the MeSHMine sheet that may be hidden and you should feel free to hide or unhide.) Here we can see a combination of MeSH supplemental concepts and MeSH heading chemicals. You can tell them apart in two ways. The MeSH supplemental concept terms start with a lower case letter (butenolide) and they are mapped to a MeSH term (4-Butyrolactone).

Most chemicals are supplemental concepts and they are mapped to MeSH heading terms. Part of the process in creating this sheet is to take each concept and behind the scenes, map it to one of its MeSH heading terms. Once a concept is mapped, it inherits tree structure and annotated subheadings from its parent.

Naturally, you'll want to go back to the Main sheet and read more about the substances you find. To do this quickly, double-click anywhere in the row (other than column A) for a chemical of interest. This action takes you to the Main sheet and fills in the first sifter cell with the substance name and sorts the rows. Double-clicking on the *sinulariolide* row gives:

	A	B	C	D	E	F	G	H	I	J
1	Abstract Sifter		Query PubMed				PubMed query run: (Anthozoa OR coral) AND (chemical toxicity OR drug effects).			Provided by the US EPA's Center for Computational Toxicology and Exposure
2	Version 8.0	Your sifter terms and frequency counts								
		sinulariolide	coral	anthozoa	Score					
3	PMID					Pub Yr	Title		Review	Authors
4	3178370	12	2	0	14	2019	Sinulariolide Inhibits Gastric Cancer Cell Migration and Invasion through Downregulation of the EMT Process			Wu YJ, Lin SH, Din ZH, Su JH, Liu CJ
5	30513611	12	3	1	16	2018	A Soft Coral-Derived Compound, 11-Dehydroisulariolide, Induces G2/M Cell Cycle Arrest and Apoptosis in S			Lin YC, Su JH, Lin SC, Chang CC, Hsia TC, Tung YT, I
6	21822415	12	1	0	13	2011	Proteomic analysis of anti-tumor effects of 11-dehydroisulariolide on CAL-27 cells.			Liu CJ, Chen CC, Chen JC, Su JH, Huang HH, Chen I
7	28767067	11	1	1	13	2017	Sinulariolide Suppresses Cell Migration and Invasion by Inhibiting Matrix Metalloproteinase-2/-9 and Urokin			Cheng TC, Din ZH, Su JH, Wu YJ, Liu CJ
8	27598175	11	1	1	13	2016	A Coral-Derived Compound Improves Functional Recovery after Spinal Cord Injury through Its Antiapoptotic			Chen CH, Chen NF, Feng CW, Cheng SY, Hung HC, I
9	26204832	11	1	0	12	2015	Sinulariolide Suppresses Human Hepatocellular Carcinoma Cell Migration and Invasion by Inhibiting Matrix M			Wu YJ, Neoh CA, Tsao CY, Su JH, Li HH
10	23973991	11	1	0	12	2013	Sinulariolide induced hepatocellular carcinoma apoptosis through activation of mitochondrial-related apopt			Chen YJ, Su JH, Tsao CY, Hung CT, Chao HH, Lin J I
11	23880933	11	1	1	13	2013	Proteomic investigation of the sinulariolide-treated melanoma cells A375: effects on the cell apoptosis thro			Li HH, Su JH, Chiu CC, Lin JJ, Yang ZY, Hwang WL, I
12	23015779	11	1	1	13	2012	Induction of apoptosis by 11-dehydroisulariolide via mitochondrial dysregulation and ER stress pathways in			Su TR, Tsai FJ, Lin JJ, Huang HH, Chiu CC, Su JH, Y I
13	23249971	9	2	1	12	2012	Induction of apoptosis by sinulariolide from soft coral through mitochondrial-related and p38MAPK pathwa			Neoh CA, Wang RY, Din ZH, Su JH, Chen YK, Tsai I I
14	22119889	9	1	1	11	2012	Neuroprotection by marine-derived compound, 11-dehydroisulariolide, in an in vitro Parkinson's model: a p			Chen WF, Chakraborty C, Sung CS, Feng CW, Jear I
15	14577690	8	2	1	11	2003	New membranoid analogues from the formosan soft coral Sinularia flexibilis and their cytotoxicity.			Hsieh PW, Chang FR, McPhail AT, Lee KH, Wu YC I
16	28901434	6	1	0	7	2017	Sinulariolide suppresses LPS-induced phenotypic and functional maturation of dendritic cells.			Chung TW, Li YR, Huang WY, Su JH, Chan HL, Lin I
17	22895288	6	2	0	8	2012	Proteomic profiling of the 11-dehydroisulariolide-treated oral carcinoma cells Ca9-22: effects on the cell			Liu CJ, Wang RY, Lin JJ, Su JH, Chiu CC, Chen JC, C
18	9827024	5	3	0	8	1998	Antimicrobial activity of the diterpenes flexibillide and sinulariolide derived from Sinularia flexibilis Quoy and			Aceret TL, Coll JC, Uchio Y, Sammarco PW
19	27801783	4	3	1	8	2016	Cytotoxicity of 11-epi-Sinulariolide Acetate Isolated from Cultured Soft Corals on HA22T Cells through the E			Lin JJ, Wang RY, Chen JC, Chiu CC, Liao MH, Wu Y I
20	26950100	4	1	1	6	2016	Anticancer Effects of Sinulariolide-Conjugated Hyaluronan Nanoparticles on Lung Adenocarcinoma Cells.			Hsiao KY, Wu YJ, Liu ZN, Chuang CW, Huang HH, I
21	11166674	4	2	0	6	2001	Discrimination between several diterpenoid compounds in feedine by Gambusia affinis.			Aceret TL, Sammarco PW, Coll JC, Uchio Y

Figure J-4. Double-clicking on Copper on the MeSHMine sheet brings you here.

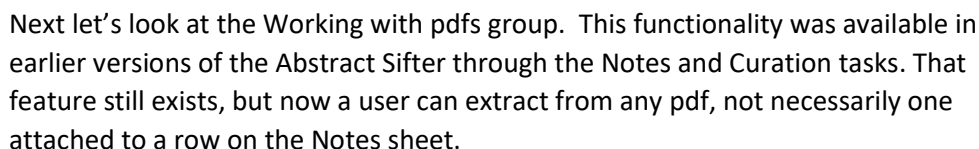
To get back to the MeSHMine sheet, just

click on the MeSHMine tab.

When you've found the substances you want to work with more, copy the names and paste them onto the Landscape sheet Column C. Or you can select rows on MeSHMine and click on the *More* button then *Send to CuratedLists*. This is a sheet used to keep a list of things of interest. It holds the query language and any other information you want to keep track of. From the CuratedLists sheet, select rows and Send to Landscape for in-depth examination.


For more in depth under-the-hood discussion of MeSH mining, see Section M.1 of this guide.

1. Button group - Working with pdfs



The Extract Tables button will let you select a pdf and (in most cases) extract the data from (some) tables. There are many variables playing into this feature and unless those variables are lined up just right, this feature may not work. Caveats aside, when it does work, it can be very helpful. Let's look at a few examples.

Click on the Extract tables button. The Abstract Sifter may ask you if you want to create a new sheet. Then it will show the file manager dialog box so you can select a pdf. In this example, we'll select the pdf for the article above.



	A	B	C	D	E	F	G
	1	PDF Extraction					
	2	pdf_text	C:\Users\Shahar\OneDrive - Environmental Protection Agency (EPA)\Profile				
	3	Table object not created but library is a table - 1 row only					
	4	Published by Oxford University press on behalf of the Society of Toxicology 2010. This work is written by EU Governors					
	5	Published by Oxford University press on behalf of the Society of Toxicology 2010. This work is written by EU Governors					
	6	Table object not created but library is a table - 1 row only					
	7	(Figure 1) Workflow for the ToxCast TSM dataset. "Samples" indicates chemicals tested in highpilot from stock plates					
	8	Table object not created but library is a table - 1 row only					
	9	Figure 2. Snapshot of highpilot test results for TSM response data (75% low/25% unknown/whether points). Top response					
	10	Table object not created but library is a table - 1 row only					
	11	Figure 3. Identification of 200 TSM positive chemicals (as identified by testagent index (TP), testability (TV), and low					
	12	response) and meta data from highpilot and check results around page 8					
	13	Section 1					
	14	1	all-chem-Biotonic acid	797		Valproic acid	160
	15	2	Pharmac_027_47333	74		Hydroxyurea	161
	16	3	Misc	3		Fluorinated	162
	17	4	Spirooxime	76		Tetracycline	163
	18	5	SAR15040	6		3,3'-Dimethylthiobis(4-methyl-5-pyridyl)	164
	19	6	7-Azido-4-hydroxybenzoic acid	76		Propargylamine	165
	20	7	Adipate 2-deoxyribose	79		Lutetium	166
	21	8	7,2-Diisobutyl-4-methyl-5-pyridyl	7		2,4-Dinitrobenzoic acid	167
	22	9	7,2-Diisobutyl-4-methyl-5-pyridyl	81		Fluorinated	168
	23	10	Carbamazepine	167		Carbamazepine	169
	24	11	Hydroxyurea	167		Hydroxyurea	170
	25	12	7,2-Diisobutyl-4-methyl-5-pyridyl	11		Fluorinated	171
	26	13	Dehydrochloride	167		Dehydrochloride	172
	27	14	Chlorophenol	167		Chlorophenol	173
	28	15	Hydroxyurea	167		Hydroxyurea	174
	29	16	Carbonyl	88		Dioxane	175
	30						
	31						
	32						
	33						
	34						
	35						
	36						
	37						
	38						
	39						
	40						
	41						
	42						
	43						
	44						
	45						
	46						
	47						
	48						
	49						
	50						
	51						
	52						
	53						
	54						
	55						
	56						
	57						
	58						
	59						
	60						
	61						
	62						
	63						
	64						
	65						
	66						
	67						
	68						
	69						
	70						
	71						
	72						
	73						
	74						
	75						
	76						
	77						
	78						
	79						
	80						
	81						
	82						
</							

a) *Extracting figures*

40



Figure I4. Images can be extracted with three size selections.

2. Button group - Get DSSTox IDs

Let's look at the way the next section on the ribbon works:



This button retrieves the EPA DSSTox database chemical identifier using either a chemical name or CAS number as input.

Here's how to use it:

Figure 15. Get DSSTox IDs button group

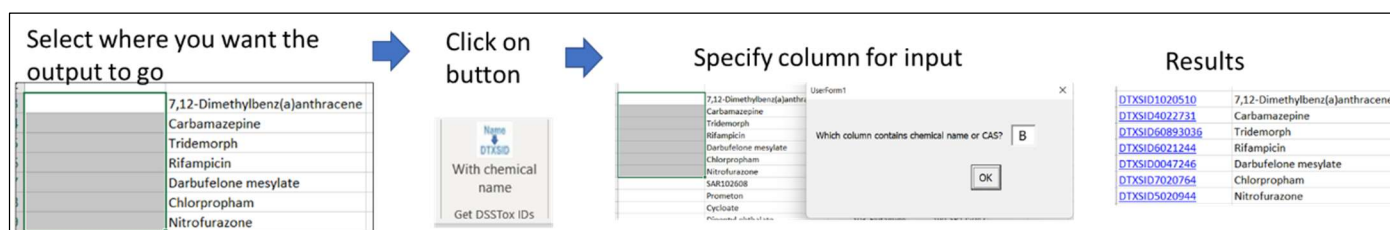


Figure 16. How to use the Get DSSTox IDs function.

If the chemical name you are pointing to is a synonym for the DSSTox identifier returned, the identifier will be shaded gray, and a message will appear.

Given the variations in chemical names and their spelling, it is sometimes a challenge to find the identifier for a name pulled from a publication. Sometimes the chemical can be found in PubChem or Common Chemistry. For information on these sources, see the PubChem section of this user guide. The synonym resolver software used by the EPA Computational Chemistry Dashboard is also useful because it will recommend spellings.

3. Button group – Get structure diagrams

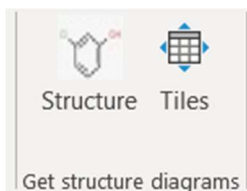


Figure K-1. Get structure diagrams ribbon group.

This ribbon group allows the user to retrieve chemical structural diagrams. These are png files delivered by the CCTE APIs.

There are two ways to do this. Let's walk through some examples.

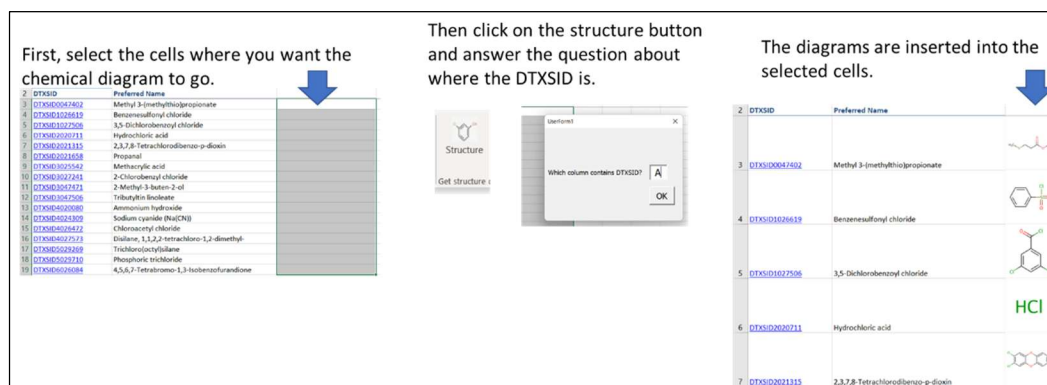


Figure K-2. Steps in retrieving chemical structure diagrams inserted into rows.

The second way to get structure diagrams is via the Tiles button.

Let's work through a basic example and a fancy example of Tiles. A list of chemicals with DTXSIDs is required. Select the rows of interest and click on the *Tiles* button. A form appears that asks for columns with input parameters. We'll ignore the colorize input and specify where the DTXSID is and the label. We're using chemical name as the label. After clicking on OK, the following structure diagrams and labels are written to the Tiles sheet. Note that the labels are in the odd rows starting with 5.

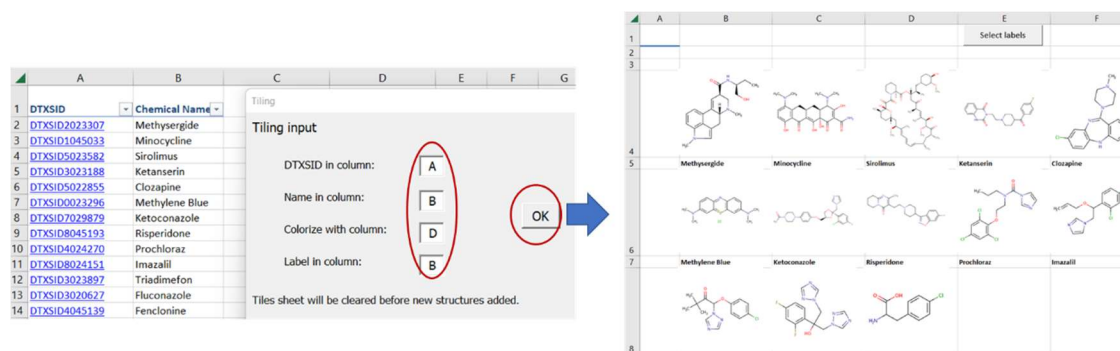

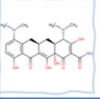



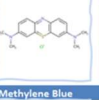
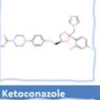
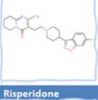
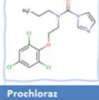
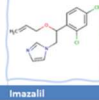
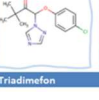
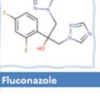



Figure K-3. Select cells and click on Tiles to create a separate sheet with a set of structure diagrams.

At this point, there are several ways to format the sheet. Clicking on the Select labels button will predictably select the labels which lets the user change the format to all at the same time. A colored frame can also be added to all the structure diagrams but clicking on one diagram, then Ctl-A and using

the Excel formatting options. Here's the output after formatting the labels and the diagrams. (Note: this in Figure K-4 is a totally fictional example.)

	A	B	C	D	E	F
1	Chemicals found in water sample					
2						
3						
4						
5	Methysergide	Minocycline	Sirolimus	Ketanserin	Clozapine	
6						
7	Methylene Blue	Ketoconazole	Risperidone	Prochloraz	Imazali	
8						
9	Triadimefon	Fluconazole	Fenclonine			

Now let's add some more information to the labels. We'll make a label that's a concatenation of various values including the chemical name, DTXSID, and the (fictional) numeric value. We'll add line feed (char(10)) to the formula to make it easier to read. (Figure K-5)

Figure K-4. Tiles can be fancied up.

DTXSID	Chemical Name	Some number
DTXSID05022855	Clozapine	96.8
DTXSID04045139	Fenclonine	-117.5
DTXSID0020627	Fluconazole	-104.0
DTXSID08024151	Imazali	-18.3
DTXSID0023188	Ketanserin	151.6
DTXSID00229879	Ketoconazole	78.6
DTXSID00023296	Methylene Blue	84.1
DTXSID0023307	Methysergide	302.4
DTXSID04045139	Minocycline	267.5
DTXSID04024270	Prochloraz	39.7
DTXSID08045193	Risperidone	77.8
DTXSID05023582	Sirolimus	233.3
DTXSID0023897	Triadimefon	-96.8

Sort and add conditional formatting

DTXSID	Chemical Name	Some number	Detailed label
DTXSID0023307	Methysergide	302.4	MethysergideDTXSID0023307Some number: 302.4
DTXSID04045139	Minocycline	267.5	MinocyclineDTXSID04045139Some number: 267.5
DTXSID05023582	Sirolimus	233.3	SirolimusDTXSID05023582Some number: 233.3
DTXSID0023188	Ketanserin	151.6	KetanserinDTXSID0023188Some number: 151.6
DTXSID05022855	Clozapine	96.8	ClozapineDTXSID05022855Some number: 96.8
DTXSID00023296	Methylene Blue	84.1	Methylene BlueDTXSID00023296Some number: 84.1
DTXSID00229879	Ketoconazole	78.6	KetoconazoleDTXSID00229879Some number: 78.6
DTXSID08045193	Risperidone	77.8	RisperidoneDTXSID08045193Some number: 77.8
DTXSID04024270	Prochloraz	39.7	ProchlorazDTXSID04024270Some number: 39.7
DTXSID08024151	Imazali	-18.3	ImazaliDTXSID08024151Some number: -18.3
DTXSID0023897	Triadimefon	-96.8	TriadimefonDTXSID0023897Some number: -96.8
DTXSID0020627	Fluconazole	-104.0	FluconazoleDTXSID0020627Some number: -104.0
DTXSID04045139	Fenclonine	-117.5	FenclonineDTXSID04045139Some number: -117.5

CHAR(10) is line feed

Sorting and adding conditional formatting to the data table.

Formula for Detailed label: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Tiling: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Colorize: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Label: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Tiling input: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Name input: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Colorize input: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Label input: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Tiling sheet: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Name sheet: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Colorize sheet: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Label sheet: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Tiling input sheet: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Name input sheet: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Colorize input sheet: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Label input sheet: `=B2 & CHAR(10) & A2 & CHAR(10) & "Some number: " & ROUND(C2,1)`

Formula for Tiling sheet cleared before new tiles added.

Figure K-5. Steps in adding more data to the Tiles.

4. Button group - Chemical names, IDs, and more

Now let's move to the along the ribbon to the group of buttons called *Chemical names, IDs, and more*.



Figure K-6. Ribbon options in the group Chemical names, IDs and more.

Here's an example showing how to retrieve CASRNs.

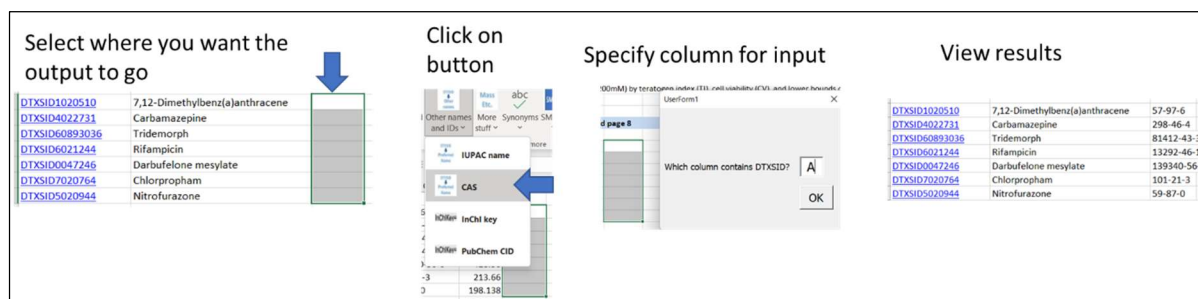


Figure K-7. Steps in running the function to retrieve CAS RNs.

The Preferred Name, Other names and IDs, More stuff, SMILES, and the PubChem buttons work in the same way.

Chemical synonyms are different from other identifiers because there can be many synonyms for one DSSTox chemical. The synonyms option in the Abstract Sifter provides two options for retrieving and viewing the synonyms. Using the figure below, let's walk through the steps. First select a row with the chemical of interest. Click on Synonyms, then See synonyms. Enter the column with the DTXSID. A form with the synonyms will appear. Synonyms have different quality descriptors such as valid and good. Scroll down to see the full list. (Figure K-8)

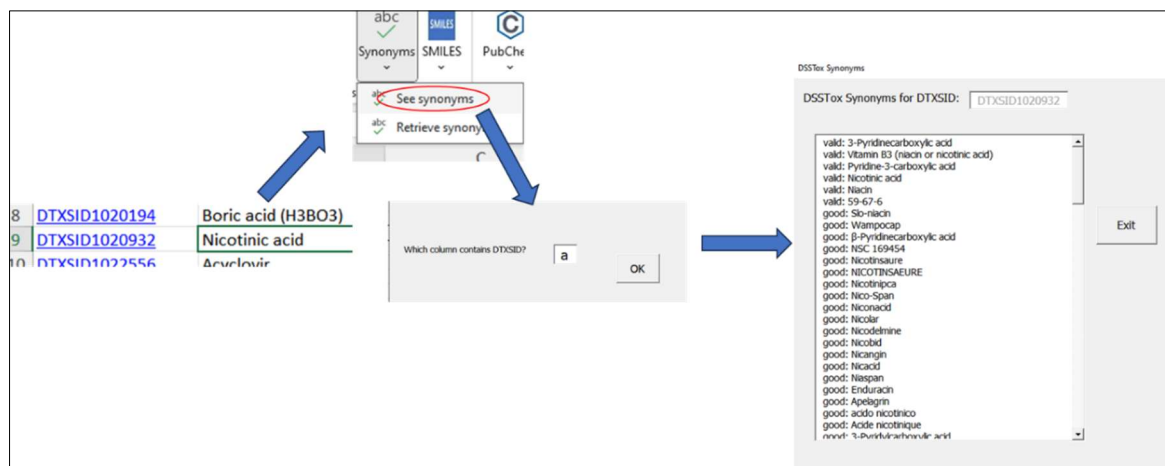


Figure K-8. Viewing synonyms for a chemical.

It's also possible to get the synonyms returned as a concatenated value in a cell. Probably the most common reason is to construct a query for a search engine like PubMed. The Abstract Sifter gives you the option, too, of formatting a query using the synonyms.

To walk through this process, select an empty cell in the row with the chemical of interest. Click on *Synonyms*, the *Retrieve synonyms*. When the synonym selection form pops up, select the types of synonyms desired, along with what delimiters and suffices to add. The selected cell is populated with the synonyms formatted as specified.

The screenshot shows an Excel spreadsheet with a table of chemicals. The 'Retrieve synonyms' dialog box is open, and the result is displayed in a cell.

DTXSID	MeSH chemical name
DTXSID0020232	Caffeine
DTXSID0021206	1,2-Propylene glycol
DTXSID0021256	Sulfasalazine
DTXSID0022519	Folic acid
DTXSID1020194	Boric acid (H3BO3)
DTXSID1020932	Nicotinic acid

The 'Retrieve synonyms' dialog box shows the following settings:

- Which column contains DTXSID? A
- Select synonym types to include:
 - ☒ Valid
 - ☒ Good
 - ☐ Alternate
 - ☐ Deleted CASRN
 - ☐ Other
 - ☐ Belstein
- Delimiter: ☐ (pipe) ☒ OR
- Add quotation marks? ☒ yes ☐ no
- Add a suffix? ☒ [tw] ☐ [all] ☐ [tiab] ☐ none
- CAS suffix? ☒ [m] ☐ none

The result in the cell is a long concatenated string of synonyms for Nicotinic acid, formatted with OR delimiters and [tw] suffixes.

Figure K-9. Writing the synonyms to a cell with concatenation and suffices.

5. Button group - PubChem and Common Chemistry



Figure K-10. PubChem and Common Chemistry.

The PubChem and Common Chemistry menus provide access to chemical data from National Library of Medicine's PubChem and American Chemical Society's Common Chemistry data resources[1, 4].

The PubChem button contains a growing number of functions that retrieve chemical entity data from PubChem and its rich APIs. (Figure K-11.)

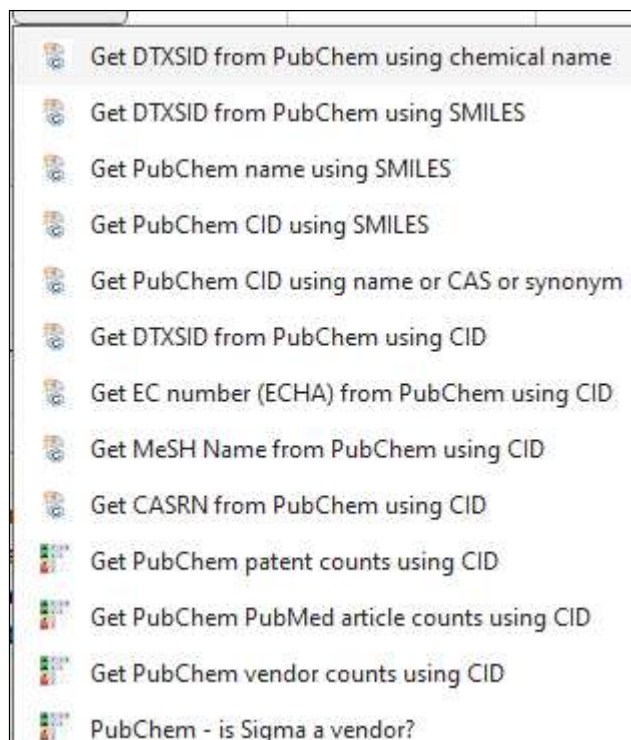


Figure K-11. PubChem function available.

The PubChem and Common Chemistry menu selections work the same way as the CCD options. The user selects an output cell, clicks on a menu option representing a function, and then specifies the column of the input data.

A list of the APIs used for each button in the PubChem menu can be found in the Technical Appendix to this user guide. The functions that retrieve the literature and patent counts do not retrieve the snapshot numbers seen on the PubChem website. The functions count the results in returned lists of patents and articles, respectively. This means that the functions are slow, especially in the case of chemicals that occur in a large number of patents. (Figure K-12) The xml returned by the request for benzene's

patent counts, for instance, will time out and give this message:

Chemical	CID	PatentCt
Benzene	cid241	time out > 200k

Figure K-12. Retrieving patent counts may time out because there are so many.

The Common Chemistry button lets the user retrieve a CAS number using chemical name.

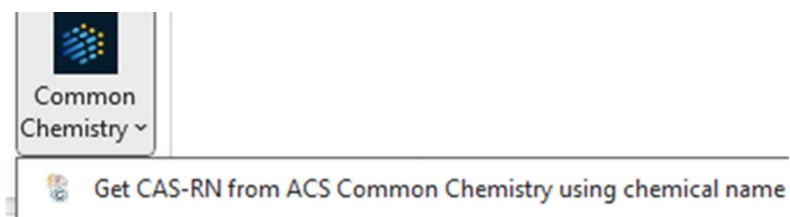
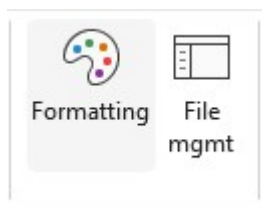


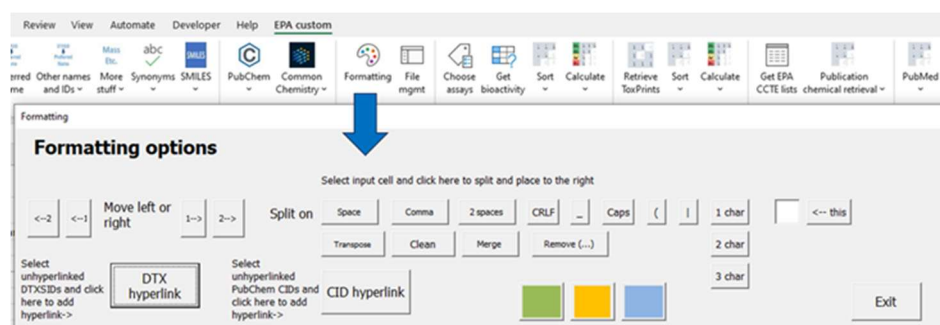
Figure K-13. Common Chemistry option.

More about Common Chemistry can be found here: [CAS Common Chemistry](#).

6. Button group – Formatting and File Management



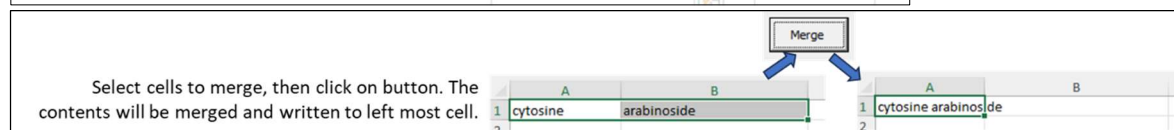
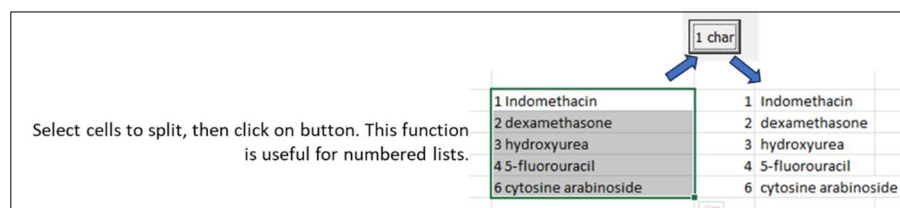
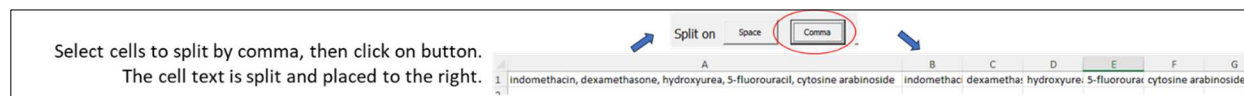
These buttons offer functionality that makes curating information from publications easier. No matter what method one uses to get information out of a pdf, the data can be messy. The formatting that makes a pdf readable does not always come across when extracting or copying and pasting. Often, too, the information that one wants is not in a nice, organized table but in a paragraph of text, and the text of interest needs to be separated from the rest of the text.



The formatting button, shown below, consists of functions that help split text, merge text, clean text, and move text. The functions work like Excel formatting functions in that you select the target cells and then click on the button. Some of the functions, like the split options, will split the target cell and write the resulting text to cells to the right of the target cell.

The move text left or right buttons will essentially cut and paste cells. You could use cut and paste as well.

Let's walk through some more of the examples.

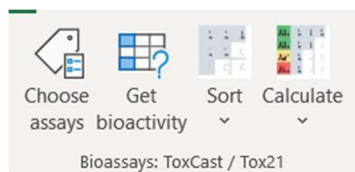


Select target cells. Click on remove. Note that the output of the function is the same as the input cell. Don't use this on chemical names with parentheses!

Sometimes DSSTox identifiers lose their hyperlinks. Use this function to add hyperlinks back to the identifiers.

The File management button lets the user point to directories. Specifying the pdf directory will direct the Abstract Sifter where to locate the downloaded and saved pdfs.

7. Button group - BioAssays: ToxCast and Tox21



This set of functions lets the end user retrieve and visualize the hit calls from EPA's ToxCast and Tox21 assays[5, 6]. Let's first get an overview of what this output might look like and then we'll go through the steps in building it.

A			B	C	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1 Bioassay Exploration					Signature:													
2 ToxCast / Tox21					Count:													
3 Beta					Query:													
4					Genes:													
5					AEID:													
6 DTXSID					ATC_UHS_CS_up													
7 DTXSID0025789					ATC_UHS_CS_up													
8 DTXSID0028038					ATC_UHS_CS_up													
9 DTXSID0032520					ATC_UHS_CS_up													
10 DTXSID0032572					ATC_UHS_CS_up													
11 DTXSID0032655					ATC_UHS_CS_up													
12 DTXSID004223					ATC_UHS_CS_up													
13 DTXSID004237					ATC_UHS_CS_up													
14 DTXSID0047034					ATC_UHS_CS_up													
15 DTXSID0047290					ATC_UHS_CS_up													
16 DTXSID0047379					ATC_UHS_CS_up													
17 DTXSID0048185					ATC_UHS_CS_up													
18 DTXSID0051493					ATC_UHS_CS_up													
19 DTXSID0051499					ATC_UHS_CS_up													
20 DTXSID0052463					ATC_UHS_CS_up													

a) Sections of the Bioassay sheet: Chemical section

Earlier parts of this document have walked through how to get a list of chemicals and DTXSIDs. Feel free to put anything in Column C and unhide columns D, E, and F. Columns from G on belong to the Abstract Sifter. This example has ToxCast active ratio in Column C. We are about to see how to retrieve ToxCast/Tox21 hit calls, so it helps to know whether a chemical actually has been tested in one of those testing programs.

b) Sections of the Bioassay sheet: Assay section

Once you have your chemicals of interest in the Chemical section with DTXSIDs in column A, then you can retrieve assays of interest and populate the Assay header section above. There are several ways to do this. Keep in mind that the ToxCast and Tox21 testing programs include over 2000 assay endpoints. Generally, people are interested in only a subset. The *Choose assays* button gives the user several ways to select a subset. The reader is encouraged to try them out. It's pretty quick.

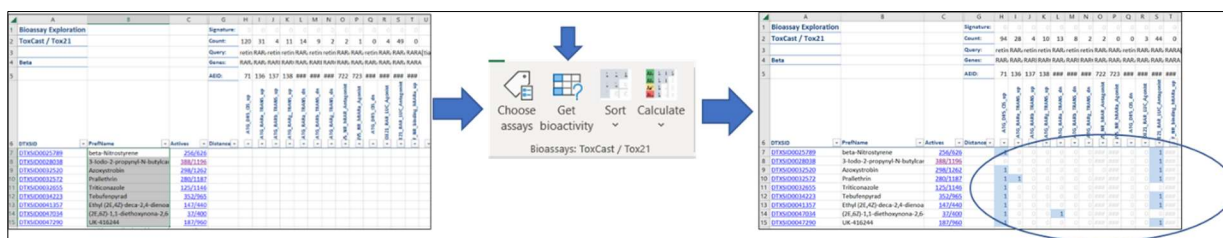
Click on the *Choose assays* button. The form below appears and presents four options for the selection of assays. In the example below the *By gene symbol* option has been selected and "rar" was entered. The assay header section was then populated with the assays. Please note that the gene search was performed by string matching and therefore is not exact entity matching. The field the match is performed against is in row 4 with Genes as the header.

The assay header section also has the assay endpoint id (AEID) in row 5, the assay endpoint name in row 6, and PubMed query terms in row 3 (more on this later).



c) Retrieving hit calls

Once the assay header information is in place, and the list of chemicals is in place, the user can retrieve the hit calls for the chemicals. To do this, select any cell in the row and click on the button *Get bioactivity*. For the chemicals in each of the selected rows, hit calls are retrieved. (Note: this can be a bit slow. A cancel option will display.)

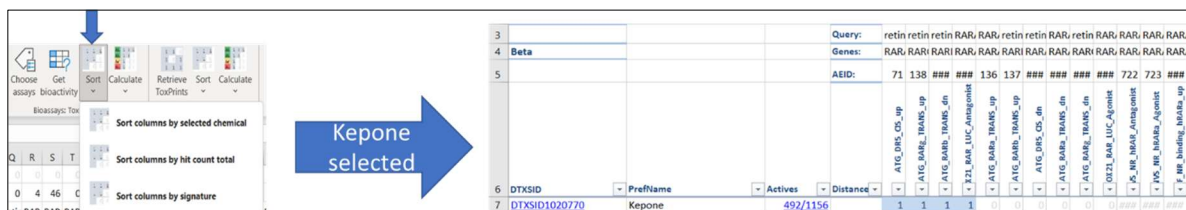


In the hit call section, a 1 means hit, 0 means not deemed a hit, and the hashes are really -9999 which means the chemical was not tested.

d) Sorting the bioassay results

Once the hit calls have been retrieved, further exploration options are available. Let's walk through the sort options available through the *Sort* button. The second sort option is straightforward – it sorts the columns by the assays with the most positive hit calls.

The first option – Sort by selected chemical – is more complicated. Let's look at it. First, select a cell in a row with a chemical of interest, then click on this option. The Abstract Sifter will move the chemical to row 7 and sort the assay columns so that positive hit calls are on the left where they are visible. That action eliminates the need for scrolling to the right to find the positive hit calls.



e) Viewing assay similarity

Visually, it is also easier to find other chemicals that might be similar to Kepone in their activity in the selected assays. If you have a large list of chemicals, manually viewing the patterns in hit calls may not be optimum so the Abstract Sifter can calculate a distance metric between your selected Row 7 chemical (Kepone in our example) and the other chemicals.

That functionality is available through the *Calculate* button. You can see by the results that Danazol assay results are closest to Kepone's.

Technical note: how is the distance calculated? The distance is always calculated between the chemical in row 7 (referred to here as Row 7 chemical) and each of the other listed chemicals (referred to here as target chemical). From an Excel perspective, the distance is calculated row by row, and for each row, column by column comparing each chemical to the Row 7 chemical. Here's the logic: The three possible values for a cell are 0, 1, or -9999 (which means not tested). If one or both of the chemicals are not tested, that column is ignored. For the rest of the cells in the row, the number of times the Row 7 value is the same as the target chemical value is counted and the number of cells that are able to be compared (that is, either 0 or 1 and not -9999) In other words the Abstract Sifter is calculating the intersection of the two sets, Row 7 chemical cells and target row cells being the two sets. The intersection is divided by the union of the two sets, giving the similarity metric.

The screenshot shows the Abstract Sifter interface with a table of results. The table has columns for DTXSID, PrefName, Actives, and Distance. The 'Calculate distance from row 7 chemical' button is highlighted with a blue arrow.

DTXSID	PrefName	Actives	Distance
DTXSID00000000	Keptone	0.000	1
DTXSID00000001	Danazol	0.000	1
DTXSID00000002	Apomorphine hydrochloride	0.000	1
DTXSID00000003	Bexarotene	0.000	1
DTXSID00000004	4-Chloro-2-cyclopentylpheno	0.000	1
DTXSID00000005	Dehydroabietamine acetate	0.000	1
DTXSID00000006	Tributyltin chloride	0.000	1
DTXSID00000007	Bromocresol	0.000	1
DTXSID00000008	beta-Nitroxyene	0.500	1
DTXSID00000009	Azoxytrobin	0.500	1
DTXSID00000010	Tetrafenylacetic acid	0.500	1

f) More on selecting assays

Let's circle back and look at the other ways to populate the Assay header section. Clicking on *By platform* reveals a selection of available assay platforms to choose from. Click on one, pick *Clear* or *Append*, and then click on *OK*. The headers will be written to the Assay Header section. Next, select chemicals and *Get bioactivity* button.

The screenshot shows the 'By platform' selection dialog with a list of platforms: ACEA, Agedica (APR), Attagene (ATG), BioSeek (BSK), CCTE, CEETOX, CLD, ENF, LTEA, and STM (STM). The 'Stemina assays' button is highlighted with a blue arrow.

DTXSID	PrefName	Actives	Distance
DTXSID00000000	Stemina	0.000	1

The option *By selected chemical* lets you retrieve all the assay headers for the positive assays for one selected chemical. So first, select a chemical. Then click on *Choose assays*, then pick *By selected chemical*.

The screenshot shows the 'Select chemical' dialog with a list of chemicals: DTXSID00000000, tert-Butylacrylamide, and DTXSID00000001, Pramlipexole dihydrochloride. The 'By selected chemical' button is highlighted with a blue arrow.

DTXSID	PrefName	Actives	Distance
DTXSID00000000	tert-Butylacrylamide	0.000	1

g) Bioassay signatures

Now let's turn to signatures. Signature is a term that refers to a set of assays that – for whatever reason – are looked at together. For instance, a biological process under study may involve a network of genes and disruption of any of these genes might perturb an important signalling network. We're calling this example gene set a signature.

The end user can specify the signature in two ways. The easiest is to go to the Signature sheet. Here all the ToxCast and Tox21 assays are listed with their genes. (Note: in the view below column D is hidden.)

Signature sheet			
Gene/Assay signature	Assay endpoint	Genes	Weight
1	AEID		
2	Assay endpoint	Genes	Weight
3	916 LTEA_HepaRG_ABCB1_dn	ABCB1	0
4	917 LTEA_HepaRG_ABCB1_up	ABCB1	0
5	1611 CLD_ABCB1_6hr	ABCB1	0
6	1627 CLD_ABCB1_24hr	ABCB1	0
7	1643 CLD_ABCB1_48hr	ABCB1	0
8	918 LTEA_HepaRG_ABCB11_dn	ABCB11	0
9	919 LTEA_HepaRG_ABCB11_up	ABCB11	0
10	1612 CLD_ABCB11_6hr	ABCB11	0

To build a signature, change the 0 in the Weight column to a 1 (or higher number to give that assay greater weight).

Sample signature			
Gene/Assay signature	Assay endpoint	Genes	Weight
1	AEID		
2	Assay endpoint	Genes	Weight
3	575 NVS_ENZ_INVEGR1	FLT1	3
4	992 LTEA_HepaRG_EGF_dn	EGF	1
5	993 LTEA_HepaRG_EGF_up	EGF	1
6	237 BSK_NDPCG_EGFR_down	EGFR	1
7	238 BSK_NDPCG_EGFR_up	EGFR	1
8	429 NVS_ENZ_NEGFR	EGFR	1
9	430 NVS_ENZ_NEGFR_Activator	EGFR	1
10	2573 ERF_ENZ_NEGFR_dn	EGFR	1
11	2885 BSK_BE3C_EGFR_up	EGFR	1
12	2886 BSK_BE3C_EGFR_down	EGFR	1
13	750 OT_Era_EREGFP_0480	ESR1	1
14	751 OT_Era_EREGFP_0480	ESR1	1
15	576 NVS_ENZ_INVEGR1_Activator	FLT1	1
16	579 NVS_ENZ_INVEGR1	FLT1	1
17	580 NVS_ENZ_INVEGR1_Activator	FLT1	1
18	183 BSK_4H_VEGFR1_down	KDR	1
19	184 BSK_4H_VEGFR1_up	KDR	1
20	577 NVS_ENZ_INVEGR2	KDR	1
21	578 NVS_ENZ_INVEGR2_Activator	KDR	1
22	916 LTEA_HepaRG_ABCB1_dn	ABCB1	0

In this sample signature, we're looking for chemicals that have activity at VEGF receptors or EGF receptors. The weights for the corresponding assays have been changed from 0 to 1 and for Vegfr1 to 3 because that gene is particularly important to the biological process under study (in our hypothetical network).

Next, we'll go back to the BioAssay sheet and click on *Choose Assays*, then pick *By signature*, then click on *OK*. The assays with the non-zero weights are written to the Assay header section.

Weight	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	FLT1	EGF	EGF	EGF	EGF	EGF	EGF	EGF	EGF	ESR1	FLT1	FLT4	FLT4	KDR	KDR	KDR	KDR	KDR	KDR	KDR
	NVS	LTEA	BSK	BSK	NVS	NVS	ERF	BSK	BSK	OT	INVS	NVS	NVS	BSK	BSK	NVS	NVS	NVS	NVS	NVS
	575	992	993	237	238	429	430	###	###	###	750	751	576	579	580	183	184	577	578	
	NVS_ENZ_INVEGR1	LTEA_HepaRG_EGF_dn	LTEA_HepaRG_EGF_up	BSK_NDPCG_EGFR_down	BSK_NDPCG_EGFR_up	NVS_ENZ_NEGFR	NVS_ENZ_INVEGR1_Activator	ERF_ENZ_NEGFR_dn	BSK_BE3C_EGFR_up	BSK_BE3C_EGFR_down	OT_Era_EREGFP_0480	OT_Era_EREGFP_0480	NVS_ENZ_INVEGR1_Activator	NVS_ENZ_INVEGR1	NVS_ENZ_INVEGR1_Activator	BSK_4H_VEGFR1_down	BSK_4H_VEGFR1_up	NVS_ENZ_INVEGR2	NVS_ENZ_INVEGR2_Activator	

Next, retrieve the bioactivity by selecting chemicals and clicking on *Get bioactivity*.

To see how the signature can help with analysis, try out two options. First sort columns by signature, then try *Calculate distance from signature* option under the *Calculate* button.

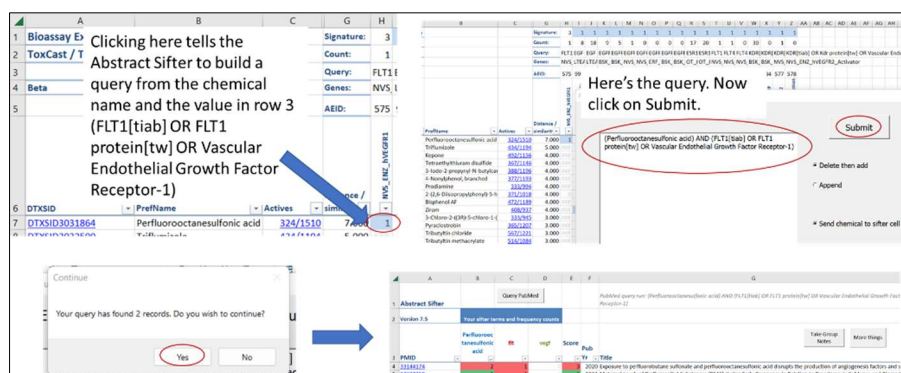
Once signature is written on to row 1, try these.

Choose assays bioactivity	Get bioactivity	Sort	Calculate	Retrieve ToxPrints	Sort	Calculate
Bioassays Tox		Sort columns by selected chemical			Sort columns by hit count total	
		Sort columns by signature				
Q	R	S	T			
1	1	1	1			
0	17					

OR

Choose assays bioactivity	Get bioactivity	Sort	Calculate	Retrieve ToxPrints	Sort	Calculate	Get lists or
Bioassays: ToxCast / Tox21							
			Calculate distance from row 7 chemical				
			Calculate distance from signature				

Let's look at the results after performing both of these actions.

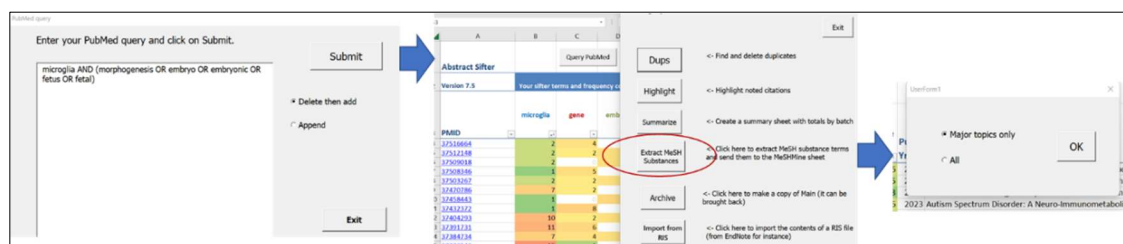


j) Populating assays from MeSHMine

One additional way to populate the assay section of the Bioassay sheets starts with the Abstract Sifter MeSHMine sheet. Let's go back and review what the MeSHMine sheet is, how it is created and how it

could be a link to assays.

The



The user can then select terms of interest – the blue colored ones - and click on the More button to have the assay header section of the Bioassay sheet populated with these assays.

Click on EPA custom to see the custom ribbon that has this ribbon group.



As we did with the BioAssays section previously, let's first get an overview of what this output might look like and then we'll go through the steps in building it.



The chemicals are placed in columns A-C. The most important element is the DTXSID because this value is used to retrieve the ToxPrints. If there is no DTXSID, then a SMILES is essential. It will be used to calculate the ToxPrints.

The Toxprints section is 729 columns with the ToxPrint numbers and names in rows 4 and 5. Row 3 contains a count of the occurrences of the ToxPrints in the chemical set.

The ToxPrint data is simply a 1 if the chemical contains the ToxPrint, 0 if not.

Columns D and E and rows 1 and 2 are useful if your chemical set has chemicals that are active or inactive in a particular assay, for example. The Abstract Sifter can help indicate which ToxPrints occur more often in the active chemicals or the inactive chemicals.

Let's go through the steps in creating this sheet.

b) Populating the chemical section

First, delete the chemicals that may already be on the sheet by selecting the rows and clicking on delete. You can also delete columns (by selecting entire columns) from column H to end. Paste your chemical names in column B and either paste or retrieve the DTXSIDs in column A. If you don't have a DTXSID for a chemical, retrieve the SMILES string in column C. (See section K above.)

DTXSID	PrefName	SMILES	EndPI	Feature enrichment score	Similarity
DTXSID00011206	1,2-Propylene glycol				
DTXSID08020541	5,5-Diphenylhydantoin				
DTXSID02020634	5-Fluorouracil				
DTXSID05021209	6-Propyl-2-thiouracil				
DTXSID02020006	Acetaminophen				
DTXSID05020027	Acrylamide				
DTXSID07021239	All-Trans-Retinoic Acid				
DTXSID07022592	Amiodarone				
DTXSID05020108	Aspirin				
DTXSID07020182	Bisphenol A				
DTXSID1020194	Boric acid				
DTXSID03020910	Busulfan				
DTXSID03020209	Butylparaben				
DTXSID00020232	Caffeine				
DTXSID04022731	Carbamazepine				
DTXSID06043709	Cyclopamine				
DTXSID05020364	Cyclophosphamide				
DTXSID03022877	Cytarabine				
DTXSID04024721	D-Camphor				
DTXSID03020384	Dexamethasone				

Figure K-16. Toxprints sheet ready to retrieve Toxprints.

c) Populating the ToxPrint headers and data sections

Next, to retrieve the ToxPrints, select rows (any cells in the rows) and then click on Retrieve ToxPrints.

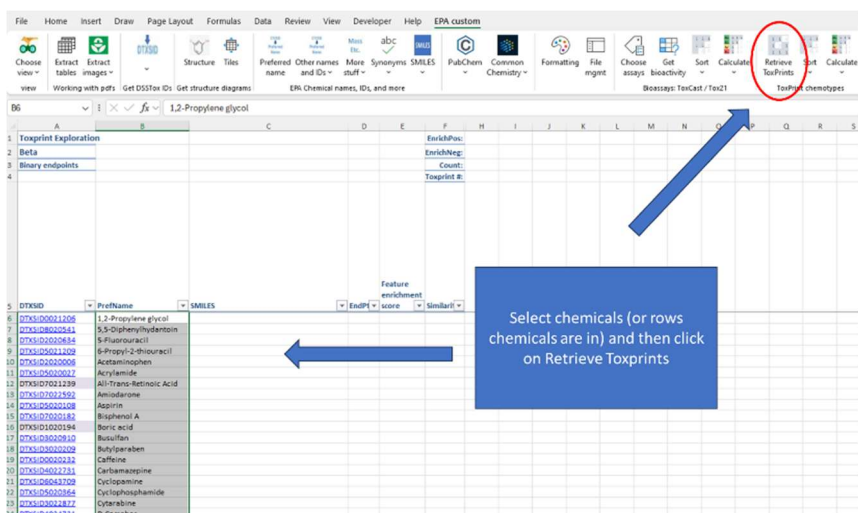


Figure K-17. To retrieve Toxprints, select chemicals and click on button.

The screenshot shows the EPA custom interface with the 'Retrieve Toxprints' button circled in red. A blue arrow points from the button to a text box that says 'Select chemicals (or rows chemicals are in) and then click on Retrieve Toxprints'. Another blue arrow points from the text box to the 'Retrieve Toxprints' button.

Figure K-18. Toxprints as 0 and 1 loaded.

The ToxPrints are returned as 0 and 1 for each chemical. Note that ToxPrints are very sparse data and at this point not exactly useful. To see which ToxPrints are associated with which chemicals, you'd have to scroll to the right a long way. Let's explore ways to make them more meaningful.

d) Sorting by selected chemical

Pick a chemical of interest, select its name, and then click on **Sort** then **Sort columns by selected chemical**.

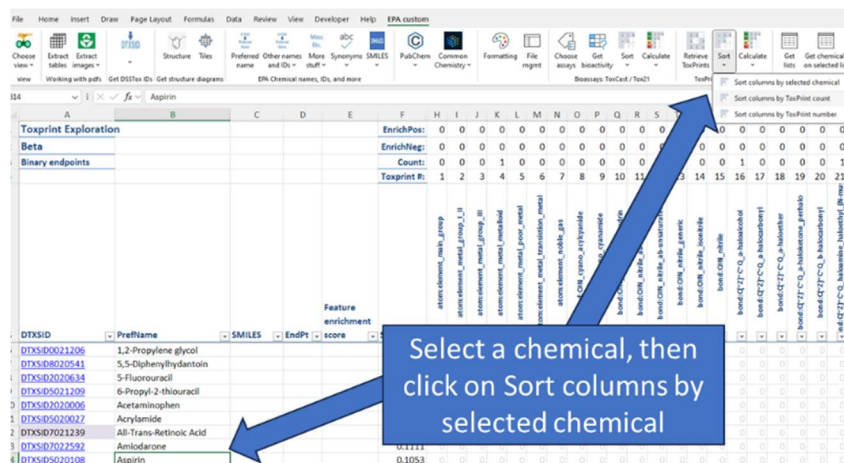


Figure K-19. Sort Toxprints by selected chemical to get a better view.

Sorting by chemical does two things. First the chemical is inserted into row 6 of the sheet. Then the columns are sorted so that the Toxprints associated with the selected chemical are visible.

DTXSID	PrefName	SMILES	EndPt	score	Similarity	EnrichPos	EnrichNeg	Count	Toxprint #
DTXSID05020108	Aspirin					0	0	0	0
DTXSID00021206	1,2-Propylene glycol					0	0	0	0
DTXSID08020541	5,5-Diphenylhydantoin					0	0	0	0
DTXSID02020634	5-Fluorouracil					0	0	0	0
DTXSID05021209	6-Propyl-2-thiouracil					0	0	0	0
DTXSID02020006	Acetaminophen					0	0	0	0
DTXSID05020027	Acrylamide					0	0	0	0
DTXSID07021239	All-Trans-Retinoic Acid					0	0	0	0
DTXSID07022592	Amiodarone					0	0	0	0
DTXSID07020182	Bisphenol A					0	0	0	0

Figure K-20. Now the Toxprints for the chemical are together and viewable. Note the selected chemical is now on row 6.

This select chemical and sort procedure can be performed as many times as wished.

e) Calculating Tanimoto distance

Often, researchers will want to pick one chemical and find out how similar the rest of the set of chemicals is to that selected chemical. To find this similarity, select a chemical, sort on it, then click on *Calculate* button, then *Tanimoto*. This action tells the Abstract Sifter to calculate the distance between the sort on chemical – the one on row 6 – and the rest of the chemicals in the set.

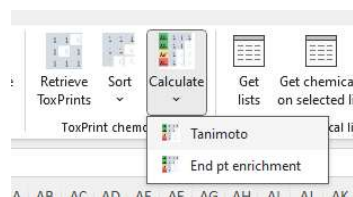


Figure K-21. To calculate the distance between the row 6 chemical and the rest of the set, click on the Tanimoto selection.

Toxprint Exploration					F	H	I	J	K	L	M	N	O	P	Q	R
Beta					EnrichPos:	0	0	0	0	0	0	0	0	0	0	0
Binary endpoints					Count:	4	9	3	3	30	15	20	0	0	0	1
					Toxprint #:	43	44	47	49	71	477	586	1	2	3	4
DTXSID	PrefName	SMILES	EndPt	score	Similarity	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic
DTXSID0502108	Aspirin			1	1	1	1	1	1	1	1	1	1	1	1	1
DTXSID07026368	Salicylic acid			0.5000	1	1	1	1	1	1	1	1	1	1	1	1
DTXSID0302455	Dimethyl phthalate			0.3750	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID02025680	MEHP			0.3333	1	1	1	1	1	1	1	1	1	1	1	1
DTXSID00021256	Sulfasalazine			0.2500	1	1	1	1	1	1	1	1	1	1	1	1
DTXSID09020740	Indomethacin			0.2222	1	1	1	1	1	1	1	1	1	1	1	1
DTXSID03020209	Butylparaben			0.2000	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID02020006	Acetaminophen			0.1538	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID06023733	Valproic Acid			0.1538	1	1	1	1	1	1	1	1	1	1	1	1
DTXSID08020541	5,5-Diphenyl			0.1429	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID05020784	Lovastatin			0.1429	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID07020182	Bisphenol A			0.1333	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID06021244	Rifampicin			0.1282	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID04022731	Carbamazepine			0.1250	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID03020465	Diethylstilbestrol			0.1250	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID00022519	Folic Acid			0.1250	1	1	1	1	1	1	1	1	1	1	1	1
DTXSID05021251	Saccharin			0.1250	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID04022949	Diphenhydramine			0.1176	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID04020822	Methotrexate			0.1176	1	1	1	1	1	1	1	1	1	1	1	1
DTXSID05023742	Warfarin			0.1111	0	0	0	0	0	0	0	0	0	0	0	0

Figure K-22. The Tanimoto distance is calculated between the Row 6 chemical and the rest of the set.

From the Tanimoto results it is evident that salicylic acid is the chemical most similar structurally in ToxPrint space to aspirin. Tanimoto is calculated with the same algorithm as the bioassay similarity described in Section K.7.

The Tanimoto can be recalculated for any number of chemicals. It's always a two-step process. First, sort by the chemical you want to calculate from.

It is inserted into row 6. Next, calculate Tanimoto.

f) End point exploration

Sometimes researchers have a set of chemicals that are associated with end points. The end points could be assay results. In the ToxPrints sheet, the researcher can explore the relationships between binary (0 or 1) end points and ToxPrints.

Let's see how that functionality works.

First, we add 0 or 1 values to the EndPoint column on the ToxPrints sheet. In this example the 0 and 1 indicate whether the chemical was positive or negative in a developmental toxicity assay[8].

Toxprint Exploration					F	H	I	J	K	L	M	N	O	P	Q	R
Beta					EnrichPos:	0	0	0	0	0	0	0	0	0	0	0
Binary endpoints					Count:	4	9	3	3	30	15	20	0	0	0	1
					Toxprint #:	43	44	47	49	71	477	586	1	2	3	4
DTXSID	PrefName	SMILES	EndPt	score	Similarity	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic	bond_C1-O2O_carboxylic_acid_aromatic
DTXSID09020740	Indomethacin		1	0.2000	1	1	1	1	1	1	1	1	1	1	1	1
DTXSID07022592	Amiodarone		1	0.1538	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID04021942	Diphenhydramine		1	0.1538	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID03020910	Buoullan		1	0.1538	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID07020182	Bisphenol A		1	0.1429	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID06021244	Rifampicin		1	0.1333	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID04022731	Carbamazepine		1	0.1282	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID02025680	MEHP		1	0.1250	1	1	1	1	1	1	1	1	1	1	1	1
DTXSID05023742	Warfarin		1	0.1250	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID06023733	Valproic Acid		1	0.1250	1	1	1	1	1	1	1	1	1	1	1	1
DTXSID08020541	5,5-Diphenylhydantoin		1	0.1250	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID00021206	1,2-Propylene glycol		0	0.1176	0	0	0	0	0	0	0	0	0	0	0	0
DTXSID02020006	Acetaminophen		0	0.1176	1	1	1	1	1	1	1	1	1	1	1	1

Figure K-23. How to calculate end point enrichment.

When the End points are in, click on Calculate then End pt enrichment. That calculates a value for each chemical and places that value in column E. Next, sort by column E in

descending order.

28	DTXSID8020541	S,S-Diphenylhydantoin	1	16.5	0.1000	0	0	0	0	0	0
29	DTXSID2025680	MEHP	1	15.9	0.0400	0	0	0	0	0	0
30	DTXSID1020194	Boric acid	1	15.4	0.0000	0	0	0	0	0	0
31	DTXSID0022519	Folic Acid	0	14.3	0.0769	0	0	0	0	0	0
32	DTXSID3020465	Diethylstilbestrol	1	12.1	0.2000	0	0	1	1	1	0
33	DTXSID5020732	Ibuprofen		7.1		0	0	0	0	0	0
34	DTXSID0020232	Caffeine	0	3.8	0.0385	0	0	0	0	0	0
35	DTXSID7020182	Bisphenol A	1	2.2	0.2105	0	0	1	1	1	0
36	DTXSID3020209	Butylparaben	0	-7.7	0.2000	0	0	1	1	1	0
37	DTXSID6025438	Hydroxyurea	1	0.0	0.0000	0	0	0	0	0	0
9	DTXSID1023819	Stavudine	1	172.5	0.1034	0	1	0	0	1	0
10	DTXSID3022877	Cytarabine	1	167.6	0.1290	0	1	0	0	1	1
11	DTXSID5020784	Lovastatin	1	164.8	0.2500	0	1	0	0	1	1
12	DTXSID3020384	Dexamethasone	1	149.5	0.0833	0	1	0	0	1	1
13	DTXSID4023753	Ziprasidone		146.2		0	0	0	0	0	0
14	DTXSID5021209	6-Propyl-2-thiouracil	1	130.2	0.0400	0	0	0	0	0	0
15	DTXSID5023742	Warfarin	1	128.6	0.2857	0	0	0	0	1	0
16	DTXSID4022949	Diphenhydramine	1	113.7	0.0417	0	0	0	0	0	0
17	DTXSID5020364	Cyclophosphamide	1	103.3	0.0400	0	0	0	0	0	0
18	DTXSID4020822	Methotrexate	1	103.3	0.0732	0	0	0	0	0	0
35	DTXSID3020209	Butylparaben		0	-7.7	0.2000	0	0	0	0	0
36	DTXSID6025438	Hydroxyurea	1	-9.9	0.0000	0	0	0	0	0	0
37	DTXSID4044292	Esomeprazole		-11.0		0	0	1	0	0	0
38	DTXSID4024721	D-Camphor	0	-25.3	0.0000	0	0	0	0	0	0
39	DTXSID5020027	Acrylamide	0	-38.5	0.0000	0	0	0	0	0	0

Figure K-26. Trying more chemicals outside the original set.

We went through the same exercise with Ziprasidone, a chemical that was positive in our assay. It ended up with structural neighbors that were all positive. Esomeprazole was tried next. This chemical, negative in the assay ended up in a mixed positive and negative neighborhood.

Predicting chemical activity using structure has a strong rationale and is a science (QSAR – quantitative structure activity relationship) with a lot of literature behind it. It's worth checking out.

Using the Abstract Sifter ToxPrints utility may indicate that structure-based methods may be a good addition to the predictive toolset.

9. Button group – Chemical lists

The Chemical lists button group offers a set of functions that are new to the Abstract Sifter. The new functionality builds on the capabilities of other public resources taking advantage of API call libraries from the EPA, PubChem, and PubMed to bring the chemicals analyzed, measured, tested, described in publications to the Abstract Sifter user.

Let's walk through them.



Figure K-27. Chemical list options.

The first button Get EPA CCTE lists is the same functionality that you can find on the EPA CCTE Computational Toxicology Dashboard (<https://comptox.epa.gov/dashboard/chemical-lists>) [6]. In that application the user picks a list and can view the chemicals.

In the Abstract Sifter, click on the *Get EPA CCTE lists* button. That action shows (or creates) a new sheet called ChemLists and populates it with the currently available list of chemical lists available through EPA CCTE DSSTox database. The list of lists can be searched using the Excel filtering and sorting options or by sifting using sifter cells B3, C3, and D3. To see more information about a list and retrieve the chemicals on it, double-click on a row with the list of interest. Double-click brings up a form with more information about the list, the number of chemicals on the list and a Retrieve button. Clicking on the *Retrieve* button causes the Abstract Sifter to create a new sheet named for the list of choice and populated with the chemicals on the list.

In the example shown in Figure I-63, the row on ChemLists with AEGLSVALUES was double-clicked and once the form appeared, *Retrieve* was clicked. This action created a sheet called AEGLVALUES populated with the 174 chemicals on the list.

ID#Name	expos	water	dust	Chemical count	Label	Description
333 40 CFR 116.4 Designation of Hazardous Substances (Above Ground Storage Tanks)						label : 40 CFR 116.4 Designation of Hazardous Substances (Above Ground Storage Tanks) Desc: i
354 40 CFR 116.4 Designation of Hazardous Substances (Above Ground Storage Tanks)						label : 40 CFR 116.4 Designation of Hazardous Substances (Above Ground Storage Tanks) Desc: i
174 AEGLS: Acute Exposure Guideline Levels						label : AEGLS: Acute Exposure Guideline Levels Desc: Acute Exposure Guideline Levels (AEGLS) i
360 CATEGORY[WIKILIST]ANTIMICROBIALS: Antimicrobials from Wikipedia						label : CATEGORY[WIKILIST]ANTIMICROBIALS: Antimicrobials from Wikipedia Desc: A list of an
401 EPA ASPECT: EPA's Airborne Spectral Photometric Environmental Collection Technology						label : EPA ASPECT: EPA's Airborne Spectral Photometric Environmental Collection Technolog
200 ATSDR: Toxic Substances Portal Chemical List						label : ATSDR: Toxic Substances Portal Chemical List Desc: The Agency for Toxic Substances and
207 ATSDR: Minimal Risk Levels (MRLs) for Hazardous Substances NAVIGATION						label : ATSDR: Minimal Risk Levels (MRLs) for Hazardous Substances NAVIGATION Desc: The Co
756 ATSDR: Minimal Risk Levels (MRLs) for Hazardous Substances (Version 1 - November 2018)						label : ATSDR: Minimal Risk Levels (MRLs) for Hazardous Substances (Version 1 - November 201
207 ATSDR: Minimal Risk Levels (MRLs) for Hazardous Substances (Version 2 - December 2022)						label : ATSDR: Minimal Risk Levels (MRLs) for Hazardous Substances (Version 2 - December 202
212 ATSDR Toxicological Profiles						label : ATSDR Toxicological Profiles Desc: Toxicological Profiles (Tox Profiles) are a unique com
726 Navigation Panel to Biosolid Lists						label : Navigation Panel to Biosolid Lists Desc: Biosolid lists change over time and are version
726 LIST: Chemicals in biosolids (2021)						label : LIST: Chemicals in biosolids (2021) Desc: Biosolids are a product of the wastewater treat
739 LIST: Chemicals in biosolids (2022)						label : LIST: Chemicals in biosolids (2022) Desc: Biosolids are a product of the wastewater treat
92 WATER EPA: Chemical Contaminants - Navigation Panel to Chemical Candidate Lists						label : WATER EPA: Chemical Contaminants - Navigation Panel to Chemical Candidate Lists Desc
50 WATER EPA: Chemical Contaminants - CCL 1						label : WATER EPA: Chemical Contaminants - CCL 1 Desc: The Contaminant Candidate List (CCL
42 WATER EPA: Chemical Contaminants - CCL 2						label : WATER EPA: Chemical Contaminants - CCL 2 Desc: The Contaminant Candidate List (CCL
106 WATER EPA: Chemical Contaminants - CCL 3						label : WATER EPA: Chemical Contaminants - CCL 3 Desc: The Contaminant Candidate List (CCL
100 WATER EPA: Chemical Contaminants - CCL 4						label : WATER EPA: Chemical Contaminants - CCL 4 Desc: The Contaminant Candidate List (CCL
93 WATER EPA: Chemical Contaminants - CCL 5						label : WATER EPA: Chemical Contaminants - CCL 5 Desc: The Contaminant Candidate List (CCL
10246 WATER EPA: Chemical Contaminants - CCL 5 PFAS subset						label : WATER EPA: Chemical Contaminants - CCL 5 PFAS subset Desc: The Contaminant Candic
8035 CDR: Chemical Data Reporting 2016						label : CDR: Chemical Data Reporting 2016 Desc: The <a href="https://www.epa.gov/chemical-i
8039 CDR: Chemical Data Reporting 2020						label : CDR: Chemical Data Reporting 2020 Desc: The <a href="https://www.epa.gov/chemical-i
5233 EPA CHEMINV: EPA Chemical Inventory for ToxCast (20170203)						label : EPA CHEMINV: EPA Chemical Inventory for ToxCast (20170203) Desc: CHEMINV consists
558 EPA CHEMINV: EPA ToxCast ChemInventory DMSO Insolubles at 20mM						label : EPA CHEMINV: EPA ToxCast ChemInventory DMSO Insolubles at 20mM Desc: Group of c
24 EPA CHEMINV: EPA ToxCast ChemInventory list of reactives						label : EPA CHEMINV: EPA ToxCast ChemInventory list of reactives Desc: ToxCast Chemical inv
34 EPA CHEMINV: EPA ToxCast ChemInventory chemicals with stability problems						label : EPA CHEMINV: EPA ToxCast ChemInventory chemicals with stability problems Desc: Toi
130 EPA CHEMINV: EPA ToxCast ChemInventory list of volatiles						label : EPA CHEMINV: EPA ToxCast ChemInventory list of volatiles Desc: List of chemicals in EPA's To
45358 Chemical and Products Database v1						label : Chemical and Products Database v1 Desc: This is a list of chemicals reported in the EPA's
391 Clean Water Act (CWA) Section 311(b)(2)(A) list						label : Clean Water Act (CWA) Section 311(b)(2)(A) list Desc: The Clean Water Act (CWA) Sectio
275 DEA Schedule 1 Drugs						label : DEA Schedule 1 Drugs Desc: Schedule I drugs, substances, or chemicals are defined as dr
58 DEA Schedule 2 Drugs						label : DEA Schedule 2 Drugs Desc: Schedule II drugs, substances, or chemicals are defined as dr
104 DEA Schedule 3 Drugs						label : DEA Schedule 3 Drugs Desc: Schedule III drugs, substances, or chemicals are defined as
81 DEA Schedule 4 Drugs						label : DEA Schedule 4 Drugs Desc: Schedule IV drugs, substances, or chemicals are defined as

Figure K-28. Double-click on a row for the desired list.

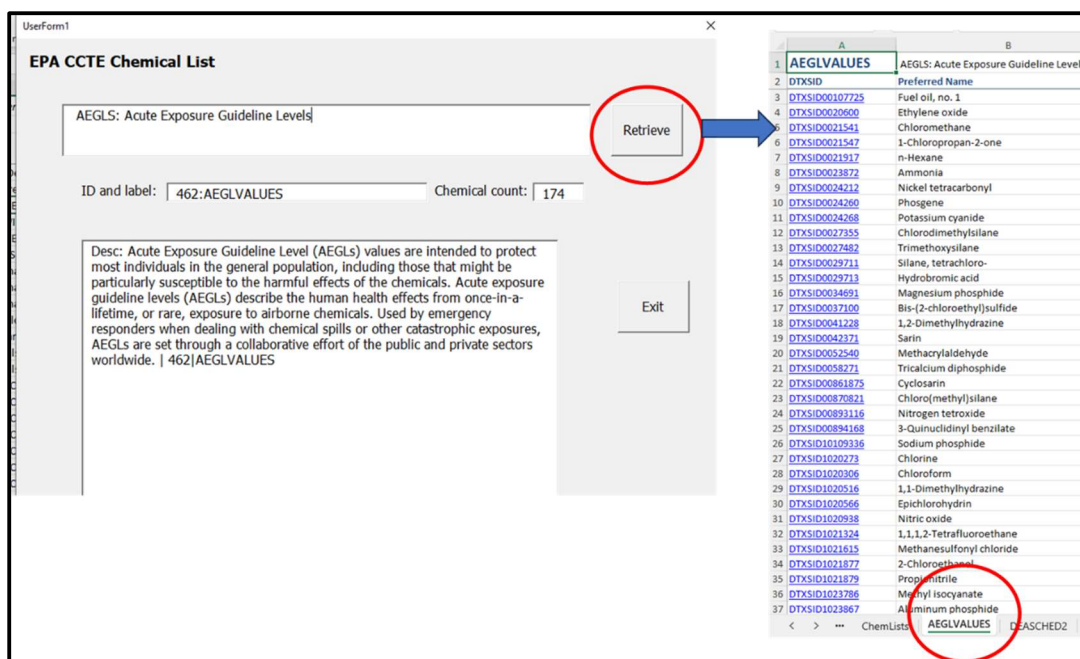


Figure K-29. Double-clicking on a list row brings up this form. Click on the Retrieve button to create a new sheet and download chemicals.

The other button under Chemical lists ribbon is Publication chemical retrieval. This button has four options as shown in Figure I-64. These options represent Excel functions that require PubMed ID as input and a place to put the output. If the user is on the Abstract sheet viewing an abstract for a PubMed record or curating a pdf, there will be a PubMed ID in cell B2. The functions will use the value in B2 as input and the output will be placed on the Abstract sheet or curation sheet, respectively.

If the user clicks on any of the options from a sheet that is not a curation or the abstract sheet, a new sheet will be created, and the user will need to enter the PubMed into cell B2.

These options are best run from either the Abstract sheet or a curation sheet for one publication. These options look for a PubMed ID in cell B2.

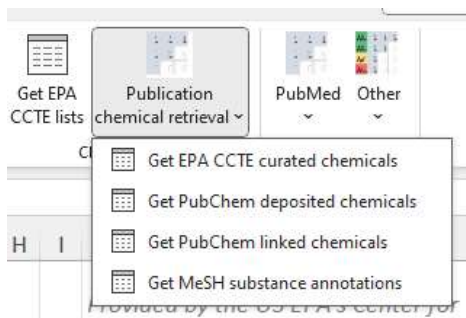


Figure K-30. The Publication chemical retrieval option has four options.

a) Option: Get EPA CCTE curated chemicals

The EPA Center for Computational Toxicology and Exposure has started a project to curate chemicals from the chemical and toxicology literature. In this context, curate means extract the chemical name and find its DSSTox identifier. The association between the document and the chemicals is stored in the cloud where they can be retrieved and presented to end users. As of this writing (April 2024) the number of articles with curated chemicals numbers only in the hundreds, but it is growing.

Other organizations also identify chemicals in publications and provide that data. The National Library of Medicine (NLM) has for many years annotated articles with a set of MeSH terms. These terms include

chemical entities. The annotation is generally restricted to the information in the abstract and lists of chemicals in the article will not likely be annotated.

PubChem provides two avenues to get chemical entities from the literature (Kim et al., 2016). First, some journals require a deposition to PubChem as a step in the publication process. These chemical entities can be retrieved from the PubMed Medline record. Second, PubChem employs automated text-mining to identify some chemical names in abstracts and links these names to the PubMed ID. For an in-depth explanation see the Kim et al. publication[1].

Chemical entities can be retrieved from any of these four sources with Abstract Sifter 8. Let's go through two examples. First, retrieve the PubMed entries for two publications: on the Main sheet click on *Query PubMed* then enter *20081854[uid] OR 33113642[uid]* into the box and click on *Submit*.

When the two entries are returned to the Main sheet, double-click on the row with 20081854, the article about zebrafish assays. This brings you to the Abstract sheet. Next, on the EPA custom ribbon, click on *Publication chemical retrieval*, then *Get PubChem deposited chemicals* (Figure I-65).

The screenshot shows the 'Abstract with highlights' sheet in Abstract Sifter 8. The ribbon includes 'EPA Chemical names, IDs, and more' and 'Publication chemical retrieval'. A blue arrow points from the 'Get PubChem deposited chemicals' button to a table of chemicals.

ID	Name	Source
85856286	isoproterenol	PubChem deposited
85856292	diazepam	PubChem deposited
85856298	apomorphine	PubChem deposited
85856299	Digitoxigenin	PubChem deposited
85856300	6-nitroquipazine	PubChem deposited
85856301	clenbuterol	PubChem deposited
85856302	Epinephrine	PubChem deposited
85856303	Dihydroergocristine	PubChem deposited
85856304	Dihydroergocristine	PubChem deposited
85856305	3,4,4a,10b-tetrahydro-4-propyl-2H,5H-(1)benzopyran(4,3-b)-2,4-dioxazin-9-ol	PubChem deposited
85856306	acetonitrile	PubChem deposited
85856318	veratridine	PubChem deposited
85856320	Strophanthidin acid	PubChem deposited
85856322	Exerine	PubChem deposited
85856314	2-(4-Azepam-1-yl)-4-oxo-butyl-1-isindole-1,3-dione	PubChem deposited
85856315	10-(2-Chlorophenyl)-1,3,4,10a-tetraaza-cyclopenta[b]fluoren-9-one	PubChem deposited
85856316	6-Oxo-7,8,9,10-tetrahydrobenzo[2]chromen-3-yl benzoate	PubChem deposited
85856309	pargoline	PubChem deposited
85856310	warfarin	PubChem deposited
85856311	aznigbas methyl	PubChem deposited
85856312	ATROPINE	PubChem deposited
85856313	naloxonium	PubChem deposited

Figure K-31. PubChem deposited chemicals for PubMed ID 20081854.

Where does this list of chemicals come from? Click on the article's PubMed hyperlink to see it on PubMed, then scroll down and click on *Associated data*. You will find the same chemicals listed here in the Abstract Sifter. The authors or the journal have deposited the chemical list with PubChem and it is surfaced through PubMed.

Next, go back to Main and double-click on the row for PubMed ID 33113642. On the Abstract sheet, pull up the *Publication chemical retrieval* button then click on *Get MeSH substance annotations* and *Get EPA CCTE curated chemicals* and see what happens. In this case there are both MeSH substances and EPA curated chemicals. (Figure I-66)

A		B			D	
		<div><div><- Main</div><div>< (shift and left arrow)</div><div>> (shift and right arrow)</div><div>Add Note</div><div>See existing notes</div><div>Like this?</div></div>				
1	Abstract with highlights					
2	PMID:	33113642			PubYr	A
3	DocID:	33113642				
4	Title:	Screening for 32 per- and polyfluoroalkyl substances (PFAS) including GenX in sludges from 43 WWTPs located in the Czech Republic - Evaluation of pote				2020
	Title and Abstract:	Screening for 32 per- and polyfluoroalkyl substances (PFAS) including GenX in sludges from 43 WWTPs located in the Czech Republic - Evaluation of potential accumulation in vegetables after application of biosolids. ABSTRACT: Highly persistent, toxic and bioaccumulative per- and polyfluoroalkyl substances (PFAS) represents a serious problem for the environment and their concentrations and fate remain largely unknown. The present study consists of a PFAS screening in sludges originating from 43 wastewater treatment plants (WWTPs) in the Czech Republic. To analyze an extended group of PFAS consisting of 32 PFAS, including GenX and other new replacements of older and restricted PFAS in sludge, a new method was optimized and validated using pressurized solvent extraction, followed by the SPE clean-up step to eliminate the observed matrix effects and LC-MS/MS. The results revealed high PFAS contamination of sewage sludge, reaching values from 5.6 to 963.2 ng g-1. The results showed that in the majority of the samples (about 60%), PFOS was the most abundant among the targeted PFAS, reaching 932.9 ng g-1. Approximately 20% of the analyzed samples contained more short-chain PFAS, suggesting the replacement of long-chain PFAS (especially restricted PFOA and PFOS). GenX was detected in 9 samples.				
5						
6	Id	Name			Source	
7	0017738	Alkanesulfonic Acids			MeSH annotated	
8	000080803	Biosolids			MeSH annotated	
9	0005308	Fertilizers			MeSH annotated	
10	0005466	Fluorocarbons			MeSH annotated	
11	0011422	Propionates			MeSH annotated	
12	0012722	Sewage			MeSH annotated	
13	0012989	Soil Pollutants			MeSH annotated	
14	0062065	Waste Water			MeSH annotated	
15	0000611729	ammonium 2,3,3,3-tetrafluoro-2-(heptafluoropropoxy)-propanoate			MeSH annotated	
16						
17	DTXSID00192353	8:2 Fluorotelomer sulfonic acid			Curated by EPA CCTE	
18	DTXSID000379268	3:3 Fluorotelomer carboxylic acid			Curated by EPA CCTE	
19	DTXSID000892447	Potassium 11-chloroicosafuoro-3-oxaundecane-1-sulfonate			Curated by EPA CCTE	
20	DTXSID10322645	N-Ethylperfluorooctane sulfonamide			Curated by EPA CCTE	
21	DTXSID10624392	2-(N-Methylperfluorooctanesulfonamido)acetic acid			Curated by EPA CCTE	
22	DTXSID1071373	N-(Methylnonafluorobutanesulfonamide			Curated by EPA CCTE	
23	DTXSID20874028	2H,2H,3H,3H-Perfluorooctanoic acid			Curated by EPA CCTE	
24	DTXSID20892484	Perfluorohexanoate			Curated by EPA CCTE	
25	DTXSID20892489	Perfluorotridecanoate			Curated by EPA CCTE	
26	DTXSID3031864	Perfluorooctanesulfonic acid			Curated by EPA CCTE	
27	DTXSID3038899	Perfluorooctanesulfonamide			Curated by EPA CCTE	
28	DTXSID40440941	N-[(Perfluorooctyl)sulfonyl]glycine			Curated by EPA CCTE	
29	DTXSID4058916	Perfluorobutanoic acid			Curated by EPA CCTE	
30	DTXSID40892481	Perfluorodecanoate			Curated by EPA CCTE	

Figure K-32. EPA curated chemicals and MeSH substances for PubMed ID 33113642.

10. Button group – Literature



Figure K-33. Literature options on ribbon.

You may be asking “Why is there a button group for PubMed when the Abstract Sifter functionality revolves around PubMed?”. Good question. We’ll go through some examples, and it will become clear.

The PubMed pull-down looks like this:

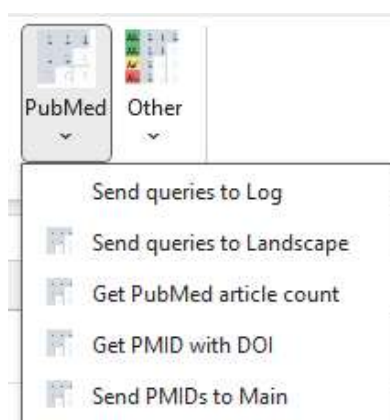


Figure K-34. PubMed options.

The first two options - Send queries to Log and Send queries to Landscape - can only be run from the BioAssay sheet. These options construct queries from the chemical names in column B and the gene identifiers in row 3 and send them either to the Log sheet or the Landscape sheet. Queries on the log

A		B	C	G	H	I	J
1	Bioassay Exploration			Signature:	0	0	0
2	ToxCast / Tox21			Count:	0	2	8
3				Query:	RAR/ RAR/ CYP2 C		
4	Beta			Genes:	RAR/ RAR/ CYP2 C		
5				AEID:	###	###	968 9
6	DTXSID	PrefName	Actives	Distance / similarity	ERF_PL_NR_Binding_hRAra	TOX21_RAR_LUC_Agonist	LTEA_HepaRG_CYP2B6
7	DTXSID7021239	Retinoic acid	239/689	###	1	1	
8	DTXSID4023177	13-cis Retinoic acid	67/253	###	1	###	
9	DTXSID6034186	Spironolactone	213/680	###	0	1	
10	DTXSID4024270	Prochloraz	306/842	###	0	1	
11	DTXSID4020119	Azathioprine	139/663	###	0	1	
12	DTXSID4020331	Desferrioxamine	14/603	###	0	1	

Figure K-35. Connecting bioassay results to the literature.

sheet can be run by double-clicking single rows or by selecting several rows and running in batch.

On the Landscape sheet the query language will be put in column C and row 3. Select the corresponding cells and click on *Update article counts* to retrieve counts.

A		B	C		D	E
Log / Batch			<div><div><-- Main</div><div>Run in batch</div></div>			
Date	Record Ct	Query Used (double-click on query to rerun)			Entry	Note
5/28/2024 9:14	0	(Azathioprine) AND (CYP2B6[tiab] OR CYP2B6 protein[tw] OR Cytochrome P-450 CYP2B6)			bioassay	
5/28/2024 9:14	0	(Prochloraz) AND (CYP2B6[tiab] OR CYP2B6 protein[tw] OR Cytochrome P-450 CYP2B6)			bioassay	
5/28/2024 9:14	0	(Spironolactone) AND (CYP2B6[tiab] OR CYP2B6 protein[tw] OR Cytochrome P-450 CYP2B6)			bioassay	
5/28/2024 9:14	1	(13-cis Retinoic acid) AND (CYP2B6[tiab] OR CYP2B6 protein[tw] OR Cytochrome P-450 CYP2B6)			bioassay	
5/28/2024 9:14	11	(Retinoic acid) AND (CYP2B6[tiab] OR CYP2B6 protein[tw] OR Cytochrome P-450 CYP2B6)			bioassay	
5/28/2024 9:14	1	(Azathioprine) AND (RARA[tiab] OR Rara protein[tw] OR Retinoic Acid Receptor alpha)			bioassay	
5/28/2024 9:14	1	(Prochloraz) AND (RARA[tiab] OR Rara protein[tw] OR Retinoic Acid Receptor alpha)			bioassay	
5/28/2024 9:14	0	(Spironolactone) AND (RARA[tiab] OR Rara protein[tw] OR Retinoic Acid Receptor alpha)			bioassay	
5/28/2024 9:14	84	(13-cis Retinoic acid) AND (RARA[tiab] OR Rara protein[tw] OR Retinoic Acid Receptor alpha)			bioassay	
5/28/2024 9:14	6503	(Retinoic acid) AND (RARA[tiab] OR Rara protein[tw] OR Retinoic Acid Receptor alpha)			bioassay	

Figure K-36. Sending to log writes the queries to the log sheet, where those not returning 0 records can be run by double-clicking.

A		B	C		E	F
Abstract Sifter		Landscape View				
Version 8			Update Article Counts	More stuff	Heat Map by column	Heat Map by row
			Query text comes from row 3 of BioAssay sheet ->		RARA[tiab] OR Rara	CYP2B6[tiab] OR CYP2B6 protein[tw]
(optional) DSSTOX link to Dashboard		Preferred Name	Subject queries: Headers come from row 6 of BioAssay sheet ->		TOX21_R	LTEA_He
			Select the cells here and click on Update Article Counts ->		AR_LUC	paRG_CY
DTXSID7021239		Retinoic acid			Agonis	P2B6
DTXSID4023177		13-cis Retinoic acid			6503	11
DTXSID6034186		Spironolactone			84	1
DTXSID4024270		Prochloraz			1	0
DTXSID4020119		Azathioprine			1	0

Figure K-37. Sending queries to Landscape populates the Landscape sheet where article counts can be retrieved.

The next selection under Literature is Get PubMed article count. Normally, the

Landscape sheet is the best place to get article count. Sometimes, though scientists want article count in a chemical list to help them rank chemicals by likelihood of being identified in water or blood by methods such as nontargeted screening.

A	B	C	D	E	F	G	H
DTXSID	Preferred Name	CAS RN	InChI key	SMILES	PubChem CID	Patent Ct	PubMed Ct
DTXSID0021096	N-Oxydiethylene-2-benzothiazylsulfenamide	102-77-2	MHKLKWCYGIBEQF-UHFFFAOYSA-N	C1CN(CCO1)SC1=NC2=C(S1)C=CC=C2	cid7619	8519	0
DTXSID1020807	2-Mercaptobenzothiazole	149-30-4	YXIWUHQXZSMYRE-UHFFFAOYSA-N	SC1=NC2=C(S1)C=CC=C2	cid697993	131883	392
DTXSID1024467	2-Aminobenzothiazole	136-95-8	UHGUUJUBCTEF-UHFFFAOYSA-N	NC1=NC2=C(S1)C=CC=C2	cid8706	14882	151
DTXSID301034849	6PPD-quinone	2754428-18-5	UBMGKRIXKUIXFF-UHFFFAOYSA-N	CC(C)CC(C)N=C1CC(=O)C(C1)=O	cid154926030	0	104
DTXSID3025178	1,3-Diphenylguanidine	102-06-7	QWRCNXZUPFZXOS-UHFFFAOYSA-N	N=C(NC1=CC=CC=C1)NC1=CC=CC=C1	cid7594	54639	136
DTXSID6061315	Benzothiazolone	934-34-9	YEDUAINPPJYJZ-UHFFFAOYSA-N	OC1=NC2=CC=CC=C2S1	cid13625	57172	132
DTXSID7024586	Benzothiazole	95-16-9	IOJULGTWVMSFF-UHFFFAOYSA-N	S1C=NC2=CC=CC=C2	cid7222	106287	17037
DTXSID9020538	N,N'-Diphenyl-p-phenylenediamine	74-31-7	UTGQNNCQVDRXCH-UHFFFAOYSA-N	N(C1=CC=CC=C1)C1=CC=C(C(NC2=CC=CC=C2)C=C1)C(C)CC(C)NC1=CC=C(C(NC2=CC=CC=C2)C=C1)C=C1	cid6319	34652	250
DTXSID9025114	6PPD	793-24-8	ZZMVLVMFYMGSMY-UHFFFAOYSA-N	CC(C)CC(C)NC1=CC=C(C(NC2=CC=CC=C2)C=C1)C=C1	cid13101	37797	139

Figure K-38. How to get article count outside Landscape sheet using Get PubMed article count.

Here's an example how to use the next options on the PubMed list of functions.

Let's say we have a list of DOIs. Maybe we got them from a bibliography. We'd like to retrieve the articles in the Abstract Sifter. First, we use the DOIs to get PubMed IDs (PMIDs), then we send the PMIDs to PubMed to retrieve the records in the Abstract Sifter.

Here's the sequence of events:

2. Click on PubMed, then Get PMID with DOI

3. Specify column containing DOI

1. Select cells for PMID

PMIDs are returned.

A	B
10.1016/j.talanta.2024.125775	38401268
10.1016/j.watres.2024.121304	38364463
10.1016/j.envpol.2024.123517	38346636
10.1007/s00204-023-03675-1	38311648
10.1007/s11356-024-32025-6	38280169
10.1016/j.envpol.2024.123364	38228259
10.1007/s11356-024-31882-5	38225497
10.1016/j.scitotenv.2023.169637	38157893
10.1007/s00204-023-03633-x	38129683
10.1016/j.scitotenv.2023.169333	38097079
10.1007/s11356-023-31281-2	38066260
10.1016/j.scitotenv.2023.169110	38065506
10.1016/j.watres.2023.120888	38039821
10.1016/j.ijbiomac.2023.127886	37926301

Figure K-39. Steps to retrieving PubMed IDs (PMIDs) with DOIs.

To get the articles sent to Main follow the steps in Figure K-27:

2. Click on Send PMIDs to Main

3. Specify column with PMIDs

4. Click on Submit

1. Select PMIDs

PubMed query

Enter your PubMed query and click on Submit.

Submit

Delete then add

Append

A	B
10.1016/j.talanta.2024.125775	38401268
10.1016/j.watres.2024.121304	38364463
10.1016/j.envpol.2024.123517	38346636
10.1007/s00204-023-03675-1	38311648
10.1007/s11356-024-32025-6	38280169
10.1016/j.envpol.2024.123364	38228259
10.1007/s11356-024-31882-5	38225497
10.1016/j.scitotenv.2023.169637	38157893
10.1007/s00204-023-03633-x	38129683
10.1016/j.scitotenv.2023.169333	38097079
10.1007/s11356-023-31281-2	38066260
10.1016/j.scitotenv.2023.169110	38065506
10.1016/j.watres.2023.120888	38039821
10.1016/j.ijbiomac.2023.127886	37926301

Figure K-40. Steps to populate Main with a list of PMIDs.

The last button on the EPA custom ribbon has (currently) only one function: it retrieves the open access link to the article's pdf, if the article is open access and the Unpaywall service has that link. It works the same as all other functions. Select the cells for output, click on the function option, specify the column for input. (Figure K-28) To find out more about the Unpaywall service, check this out:

<https://unpaywall.org/>

	A	B	C
1	10.1016/j.talanta.2024.125775	38401268	Not OA or NA
2	10.1016/j.watres.2024.121304	38364463	Not OA or NA
3	10.1016/j.envpol.2024.123517	38346636	Not OA or NA
4	10.1007/s00204-023-03675-1	38311648	web link
5	10.1007/s11356-024-32025-6	38280169	web link
6	10.1016/j.envpol.2024.123364	38228259	Not OA or NA
7	10.1007/s11356-024-31882-5	38225497	Not OA or NA
8	10.1016/j.scitotenv.2023.169637	38157893	Not OA or NA
9	10.1007/s00204-023-03633-x	38129683	web link
10	10.1016/j.scitotenv.2023.169333	38097079	Not OA or NA
11	10.1007/s11356-023-31281-2	38066260	Not OA or NA
12	10.1016/j.scitotenv.2023.169110	38065506	Not OA or NA
13	10.1016/j.watres.2023.120888	38039821	Not OA or NA
14	10.1016/j.ijbiomac.2023.127886	37926301	Not OA or NA

Figure K-41. Retrieving a link to open access pdfs using DOIs and the unpaywall service.

L. Helpful Tips and Guidelines

1. Tip 1 – Checking quality of results

The Landscape sheet is a great way to explore a set of chemicals, but some chemical names are long, complex, and a challenge to PubMed. If you copy and paste a chemical name from another source, make sure it does not have any special characters. Nonprinting escape characters make the web service calls give unexpected results, but the PubMed web site knows how to ignore them. Also, try your chemical names with quotation marks and without them to see which works better. Sometimes quotation marks are essential.

For other types of errors or strange results, it's a good idea to check it in PubMed. You can take any query generated by the Abstract Sifter and copy and paste it into PubMed using Ctl-C to copy and Ctl-V to paste. For example, the query in the box shown in Figure L-1 is selected and copied (with Ctl-C). Then in PubMed the query is pasted into the query line at the top as shown in Figure L-1. Click on the search to run the query then click on the Advanced link to be taken to the page with search details. Figure L-1 shows the information provided by PubMed about how it expands the query. If you need to learn more about PubMed queries, click on Help on the PubMed home page.

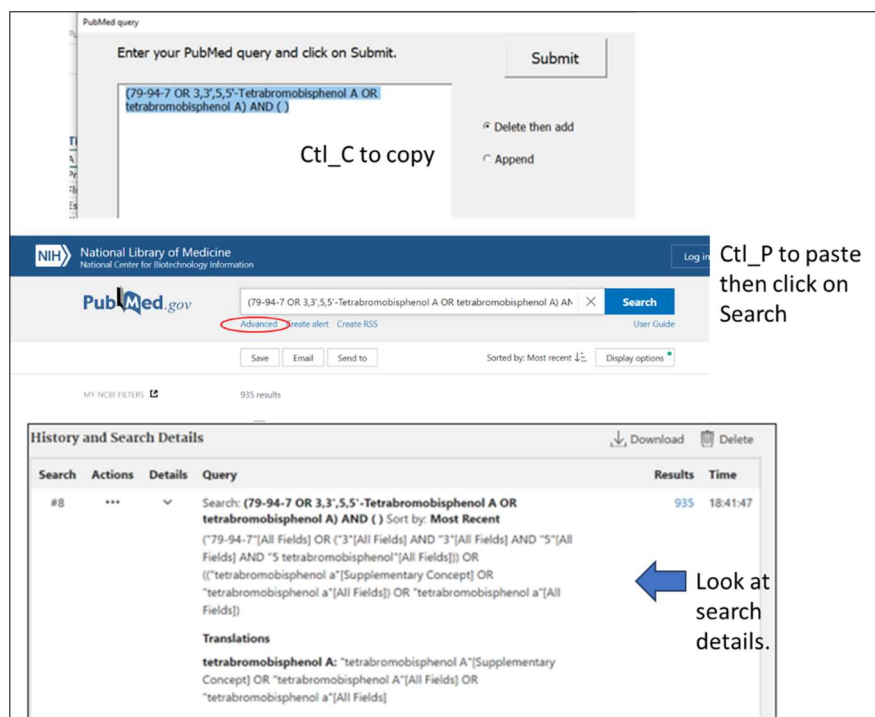


Figure L-1. How to check out a query by running it in PubMed and seeing how PubMed parses and expands the terms.

2. Tip 2 – Sifting the chemical literature

It can be very helpful in chemical research to include the chemical name in the sifting process. This is because a chemical can be mentioned in an abstract even in cases where the article is not really about the chemical and will be retrieved in the PubMed query (depending on how the query is worded).

Counting the occurrences of the chemical name in the abstract through the sifting process can help the user discriminate between articles mentioning a chemical or those that are actually about the chemical.

3. Tip 3 – Cleanup and customization

The Abstract Sifter can be cleaned up by deleting rows and columns from previous work, but the Abstract Sifter programming requires certain columns and rows to be in certain places. To learn how to clean up your sifter without disrupting the behind-the-scenes coding, consult the table below.

Sheet name	Advice for cleaning
Main	Do not add columns. Rows will be added and deleted by the Sifter.
Abstract	The Abstract Sifter rewrites the data on this sheet every time you double-click on a Main sheet row. If you want to save chemicals retrieved to the Abstract sheet (see K.9) then copy and paste them to another sheet.
Notes	Delete any unwanted rows after Row 2. Do not add or delete columns.
Log	Delete any unwanted rows after Row 2. Do not delete columns. Add columns after G if desired.
Sample_Queries	Delete unwanted rows after Row 3. Modify and add rows as desired, following the pattern of current rows. (That is, keep the heading in column B and the query text in column C.) <i>Hint: use this sheet to keep queries important to your organization.</i>
Pathway_queries	Delete unwanted rows after Row 3. Modify and add rows as desired, following the pattern of current rows.
Landscape	Delete or modify rows after Row 4 and columns after Column D.
CuratedLists	Delete or modify rows after Row 3 and columns after Column C.
TermMap	Delete or modify rows after Row 3.
MeSHMine	Abstract Sifter deletes from this sheet before writing new data. Feel free to make a copy of it.
Toxprints	Delete rows after row 5. It's OK to delete columns from H on.
Bioassay	Delete rows after row 6. OK but not necessary to delete columns from H to the right.
Tiles	Tiles is cleared before using by the Abstract Sifter.
Signature	Delete after row 2 ... but there's really no need to.

Keep in mind that the Abstract Sifter is an Excel file. You can rename it, mail it, and of course, if you want to keep your Log, Notes, and Landscape entries, you should save it. The Sample_queries sheet

provides an opportunity for you and your organization to start collecting and organizing queries that you have found useful.

4. Tip 4 – Collaborative literature review tips

Sometimes more than one person will want to work together on evaluating a set of articles. The Abstract Sifter has some features to make this easier. So, let's say Mary and Joe each retrieve, sift, and take notes on their own Abstract Sifter files. Mary can copy Joe's notes to her version of the Sifter and then she has both sets. (Or they can mail the Sifter back and forth.)

The screenshot shows the 'My Notes' section of the Abstract Sifter. A table lists notes with columns for PMID, yes/no/maybe status, Who, Tag, and Note. The 'Notes export' sidebar is open, showing options for exporting notes. The 'Highlight conflicting' button is circled in red.

PMID	yes	no	maybe	Who	Tag	Note
9860498	0	0	1	joe	Somewhat interesting	
9860498	1	0	0	mary	Helpful	
12442503	1	0	0	mary	Helpful	
14643964	0	1	0	mary		
22045597	0	0	1	joe		
22045597	1	0	0	mary	Helpful	
24978116	1	0	0	mary	Helpful	
25481984	0	1	0	mary	Do not include	
26142839	0	1	0	mary		
26142839	1	0	0	mary	Helpful	
27371222	1	0	0	mary	Helpful	
27614034	0	0	1	joe		
29881965	0	1	0	mary		
30387063	0	1	0	mary		

Figure J5. Notes from two reviewers combined on the Notes sheet.

To see if she and Joe disagreed on any record, she can click on More Stuff then Highlight Conflicting. The titles of Notes with different yes/no/maybe designations are colorized in purple. Resolving the conflicts and re-clicking on the button will cause the purple to disappear.

The screenshot shows the 'My Notes' section of the Abstract Sifter. A table lists notes with columns for PMID, yes/no/maybe status, Who, Tag, Note, PubYr, Title, and Auth. The title of the note for PMID 22045597 is highlighted in purple.

PMID	yes	no	maybe	Who	Tag	Note	PubYr	Title	Auth
22045597	0	0	1	joe			2012	Adsorption, transport and degradation of fipronil termiticide in three Haw Shua	
22045597	1	0	0	mary	Helpful		2012	Adsorption, transport and degradation of fipronil termiticide in three Haw Shua	
26142839	0	1	0	mary			2015	Fipronil induces CYP isoforms in rats.	Caba
26142839	1	0	0	mary	Helpful		2015	Fipronil induces CYP isoforms in rats.	Caba
9860498	0	0	1	joe	Somewhat interesting		1998	Mechanisms for selective toxicity of fipronil insecticide and its sulfone me	Hain
9860498	1	0	0	mary	Helpful		1998	Mechanisms for selective toxicity of fipronil insecticide and its sulfone me	Hain
12442503	1	0	0	mary	Helpful		2003	Fipronil: environmental fate, ecotoxicology, and human health concerns.	Tingl

Figure J6. Purple highlighting on title of conflicting notes.

5. Tip 5 – Connections to the EPA Chemicals Dashboard

The Environmental Protection Agency's Chemicals Dashboard is a great place to find chemical information to enhance your chemical search queries with synonyms and CAS numbers. Future releases

of the Dashboard will offer opportunities to download a list of chemicals formatted for easy insertion into the Landscape sheet. You'll find the Chemistry Dashboard here:
<https://comptox.epa.gov/dashboard>.

The EPA Chemicals Dashboard also contains its own (slightly different) version of the Abstract Sifter. It works on the same basic premise as the Excel version, but has some interesting differences. To see it, start with a chemical search. Let's look at the chemical fipronil by entering the name in the search box and clicking on the search icon (magnifying glass) (Figure J7).

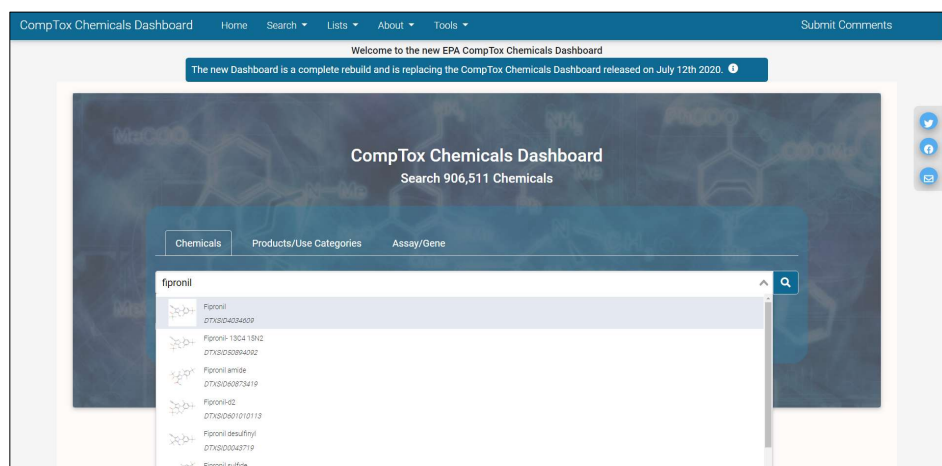


Figure J7. Searching for fipronil on the EPA's Chemistry Dashboard entry form

The main page for fipronil is displayed with the structure diagram and a selection of tabs below that lead to other information about the chemical. Click on the Literature tab as shown in Figure J8.

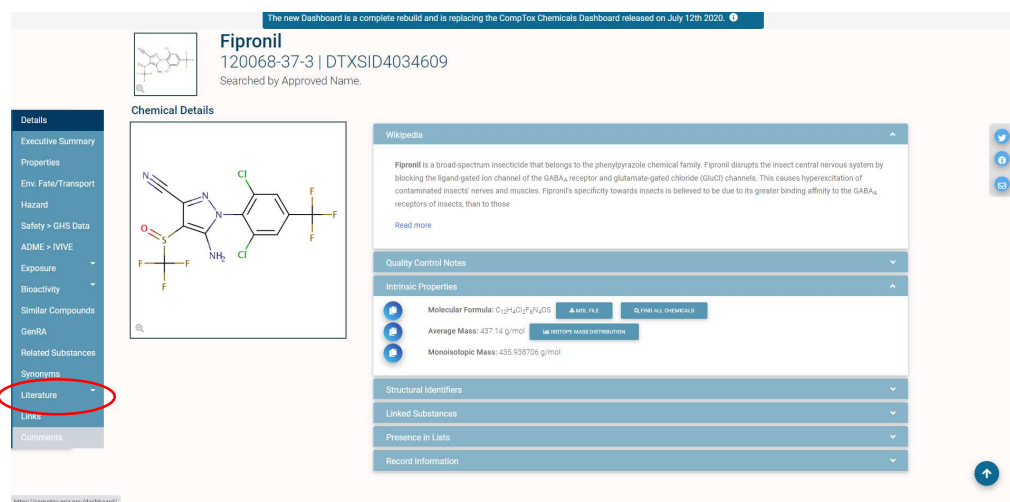
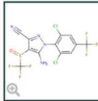


Figure J8. Select the Literature tab then on PubMed Abstract Sifter (see below).

Welcome to the new EPA CompTox Chemicals Dashboard

The new Dashboard is a complete rebuild and is replacing the CompTox Chemicals Dashboard released on July 12th 2020. [i](#)



Fipronil

120068-37-3 | DTXSID4034609

Searched by Approved Name.

Literature - PubMed Abstract Sifter

Abstract Sifter Instructions

- Select PubMed starting point query
- Optionally, enter any PubMed query or edit the query from step 1
- Click Retrieve Articles to begin download.
- Optionally, export articles

Metabolism/PK/PD


("120068-37-3" OR "Fipronil") AND (metabolism OR metabolite OR tissue distribution OR pharmacokinetics OR pharmacodynamics)

RETRIEVE ARTICLES

SEND TO

Figure J9. How to select prepared queries.

Select PubMed Abstract Sifter on the left set of buttons. The Dashboard helps you to build queries for this chemical. The chemical identifier part of the query is prepopulated on the right with name and CAS number. The subject matter part of the query is determined by selecting a topic area in the pull-down box in the center of the form. The user has several pre-composed queries to choose from. When one of them is chosen, the query is modified by appending the subject matter text. Figure J9 shows that when Metabolism/PK/PD is chosen, the text (metabolism OR metabolite OR tissue distribution OR pharmacokinetics OR pharmacodynamics) is appended to the chemical identifiers. The query can be modified manually as well. When ready, the user clicks on Retrieve Articles. See Figure J10.



Fipronil

120068-37-3 | DTXSID4034609

Searched by Approved Name.

Literature - PubMed Abstract Sifter

Abstract Sifter Instructions

- Select PubMed starting point query
- Optionally, enter any PubMed query or edit the query from step 1
- Click Retrieve Articles to begin download.
- Optionally, export articles

Metabolism/PK/PD

("120068-37-3" OR "Fipronil") AND (metabolism OR metabolite OR tissue distribution OR pharmacokinetics OR pharmacodynamics)

RETRIEVE ARTICLES

SEND TO

1073 of 1073 articles loaded

To find articles quickly, enter terms and press [RETURN] to sift abstracts.

PubMed ID	Year	Title	Authors	Journal	Rev	DOI	Vol	Issue	Pages
35934087	2022	Quantitative exposure assessment and risk characterization for fipronil...	Canton Sigrin; Canton Dominguez; Rialas; Alvarez; Lanusse...	Journal of food science		10.1111/1750-3641.15161	-	-	-
35500697	2022	Fipronil and fipronil sulfone in chicken: From in vitro experiments...	Lauze; Stoop; Giring; Hoogenboom; Punt	Food and chemical toxicology: an international journal published...		10.1016/j.fct.2022.1130...	165	-	113086
35468410	2022	Mating enhances internal concentrations of fipronil and thereby...	Hendriks; To; Takashima; Somiya; Taki; Oshima; Ohubo	Aquatic toxicology (Amsterdam, Netherlands)		10.1016/j.aquatox.2022...	247	-	109172
35410476	2022	Identification and validation of ATP-binding Cassette Transporters...	Guan; Yang; Jiang; Zhang; Wu; Jiang; Shen; Qian; Wang; Meng	Journal of agricultural and food chemistry		10.1021/acs.jafc.2c00474	70	15	4871-4819
35357189	2022	Occurrence of Phenylpyrazole and Diamide insecticides in Lactari...	Ull; Chen; Luy; Wu; Li; Zhao; Wu	Journal of agricultural and food chemistry		10.1021/acs.jafc.2c008...	70	14	4467-4474
35334479	2022	Dietary Exposure to Bifenthrin and Fipronil Impacts Swimming Re...	Magnuson; Fuller; Huff; Hartz; Anzalone; Whitledge; Aouf; Lydy...	Environmental science & technology		10.1021/acs.est.1c06609	56	8	5071-5080
35342797	2022	In-vivo and in-vitro effectiveness of three insecticides types for en...	Abouelela; Sobhan; Abouelhasan; Farid; Soliman	Open veterinary journal		10.5455/OVJ.2022.12...	12	1	44-60
35191025	2022	Enantioselective toxicity, degradation and transformation of the c...	Ou; Yan; Shi; Yu; Cai; Ma	Ecotoxicology and environmental safety		10.1016/j.ecoenv.2022...	236	-	113424
35279553	2022	Development and validation of software that quantifies the animal...	Sousa; Figueiredo; Tavares; Gomes; Kalfas; Silva; Costa-Junior...	Ticks and tick-borne diseases		10.1016/j.tiddis.2022.1...	13	3	101930
35176695	2022	Chronic environmentally relevant levels of pesticides disrupt ener...	Pinto; Rocha; Moreira; da Silva; Toshi; Goulart; Montagner; Daam...	Aquatic toxicology (Amsterdam, Netherlands)		10.1016/j.aquatox.2022...	245	-	108117
35198446	2022	Irrigation with Water Contaminated by Sugarcane Pesticides and ...	Ogura; Moreira; da Silva; Nogueira; Freitas; da Silva Pinto; Lopes; Yo...	Archives of environmental contamination and toxicology		10.1007/s00244-022-0...	82	3	330-340

J10. Results after retrieving articles.

Abstract Sifter Instructions

1 Select PubMed starting point query
Metabolism/PK/PD

2 Optionally, enter any PubMed query or edit the query from step 1
("120068-37-3" OR "Fipronil") AND (metabolism OR metabolite OR tissue distribution OR pharmacokinetics OR pharmacodynamics)

3 Click on Retrieve Articles

To find articles quickly, enter terms and press [RETURN] to sift abstracts.

fipronil tissue metabol

← Sifter terms

fipronil	tissue	metabol	Total	PubMed ID	Year	Title	Authors	Journal
14	0	3	17	35534087	2022	Quantitative exposure assess...	Canton; Signorini; Canton; Do...	Journal of food science
9	0	4	13			... sulfone in...	Lautz; Stoopen; Ginting; Hoo...	Food and chemical toxicol...
11	0	2	13			... internal co...	Hano; Ito; Ito; Takashima; So...	Aquatic toxicology (Amster...
1	1	1	3	35410476	2022	Identification and Validation o...	Guan; Yang; Jiang; Zhang; Wu...	Journal of agricultural and
4	0	1	5	3557189	2022	Occurrence of Phenylpyrazol...	Liu; Chen; Yu; Wu; Li; Zhao...	Journal of agricultural and
7	0	1	8	3553479	2022	Dietary Exposure to Bifenthrin...	Magnuson; Fuller; Huff Hartz...	Environmental science & te...
4	0	0	4	35342737	2022	In-vivo and in-vitro effectiven...	Abouelela; Sobieh; Abouelhas...	Open veterinary journal
12	0	4	16	35313125	2022	Enantioselective toxicity, degr...	Ou; Yan; Shi; Yu; Cai; Ma	Ecotoxicology and environm...
2	0	0	2	35279553	2022	Development and validation ...	Sousa; Figueiredo; Tavares; Go...	Ticks and tick-borne diseas...
5	0	0	5	35176695	2022	Chronic environmentally relev...	Pinto; Rocha; Moreira; da Silv...	Aquatic toxicology (Amster...

J11. Sifting and sorting.

Figure J10. Sifting on the EPA Chemicals Dashboard's PubMed Abstract Sifter.

After the user clicks on Retrieve Articles, the article information is retrieved from PubMed and inserted into the results table. The articles can be sifted by entering terms into the boxes shown. In the example in Figure J10, the user has entered "fipronil", "tissue", and "metabol" into the sifter entry boxes. The occurrences of these terms are counted for each PubMed citation and displayed. The table can be sorted on these values. Clicking on a row tells the Dashboard Sifter to display the title and abstract below the table with the sifter terms highlighted as in Figure J11.

To find articles quickly, enter terms and press [RETURN] to sift abstracts.

fipronil tissue metabol

← Sifter terms

fipronil	tissue	metabol	Total	PubMed ID	Year	Title	Authors	Journal	DOI	Vol	Issue	Page
14	0	3	17	35534087	2022	Quantitative exposure assess...	Canton; Signorini; Canton; Do...	Journal of food science	10.1002/food.12345	74	13	4611...
9	0	4	13	3557189	2022	Occurrence of Phenylpyrazol...	Liu; Chen; Yu; Wu; Li; Zhao...	Journal of agricultural and foo...	10.1021/acs.jafc.3c00123	70	14	448...
11	0	2	13	3553479	2022	Dietary Exposure to Bifenthrin...	Magnuson; Fuller; Huff Hartz...	Environmental science & tech...	10.1021/acs.est.3c00123	56	8	507...
4	0	0	4	35342737	2022	In-vivo and in-vitro effectiven...	Abouelela; Sobieh; Abouelhas...	Open veterinary journal	10.5455/ovj.2022.12.1.44...	12	1	44...
12	0	4	16	35313125	2022	Enantioselective toxicity, degr...	Ou; Yan; Shi; Yu; Cai; Ma	Ecotoxicology and environme...	10.1016/j.ecoenv.2022.1134...	235	-	1134...
2	0	0	2	35279553	2022	Development and validation ...	Sousa; Figueiredo; Tavares; Go...	Ticks and tick-borne diseases	10.1016/j.tiddis.2022.1019...	13	3	1019...
5	0	0	5	35176695	2022	Chronic environmentally relev...	Pinto; Rocha; Moreira; da Silv...	Aquatic toxicology (Amsterda...	10.1016/j.aquatox.2022.1081...	245	-	1081...
2	0	0	2	35198448	2022	Irrigation with Water Contami...	Ogura; Moreira; da Silva; Neg...	Archives of environmental co...	10.1007/s10646-022-02330-...	82	3	330...
4	0	0	4	35118512	2022	Effect of the formulation with ...	Secchi; Vale; de Castro Rodri...	Parasitology research	10.1007/s00336-022-02330-...	121	3	839...
12	8	1	21	35085889	2022	Tissue-specific accumulation, ...	Wang; Cheng; Pan; Luo	Ecotoxicology and environme...	10.1016/j.ecoenv.2022.1132...	232	-	1132...
7	0	1	8	35085654	2022	Realistic exposure to fipronil...	Silberschmidt Freitas; da Silva	Environmental pollution (Bari...	10.1016/j.envpol.2022.1188...	299	-	1188...
1	0	0	1	35002766	2022	Laboratory and Field Evaluati...	Yasuda; Matsubara; Hsu; Lee...	Journal of economic entomol...	10.1093/jee/taab001	115	2	624...

Tissue-specific accumulation, transformation, and depuration of fipronil in adult crucian carp (Carassius auratus).
Accumulation and biotransformation of pesticides in fish (Silberschmidt et al., 2022) are essential to assess their toxicity and associated human exposure risk. The mechanisms on time-dependent and tissue-specific accumulation and transformation of fipronil in adult fish are limited. An experiment consisting of 25-d uptake of fipronil at two levels (1.0 and 50 µg/L) and 25-d depuration in adult crucian carp (Carassius auratus) was conducted. Fipronil concentration at 25-d exposure was tissue-specific with the order of liver > kidney > blood > muscle. The uptake rate constant of fipronil in the liver (low exposure group: 2.38 ± 0.27 L/kg/d; high exposure group: 1.10 ± 0.11 L/kg/d) was significantly higher than that in other tissues (p < 0.05), and the lowest in muscle (low exposure group: 0.10 ± 0.01 L/kg/d; high exposure group: 0.16 ± 0.11 L/kg/d). The bioconcentration factors of fipronil in different tissues were 1.04-12.7 L/kg wet weight and 177-4268 L/kg lipid. The tissue-blood distribution coefficients of the liver and kidney were lower than 1 based on lipid normalized concentration but higher than 1 based on wet weight concentration, suggesting fipronil was dispersed into other tissues mainly via blood in the lipid-combination pattern. Fipronil sulfone had 1.2-32 times higher concentration and longer depuration time than fipronil, implying fipronil sulfone was more retender in fish bodies. The estimated daily intake of fipronil via fish muscle consumption at 25-d exposure was 8.5-101 and 27-320 ng/kg bw/d for adults and children, respectively. Overall, the human health risk of fipronil and its metabolites with consumption of the polluted fish cannot be negligible.

Figure J11. Clicking on a row displays the highlighted abstract below.

A check box on the left of the table provides a way to select citation rows. Selected rows can be downloaded or sent to PubMed by clicking on the pull-down box to the right.

6. Tip 6 - Populating the Excel Abstract Sifter from the EPA CompTox Chemicals Dashboard

The Chemicals Dashboard can download chemicals in a variety of formats. One of those formats make it easy to use in the Excel Abstract Sifter. Here's an example to get you started. On the home page of the Dashboard, click on Lists, then chemical, then pick a list. We'll pick Algal Toxins as a sample. Click on the name, then, when the chemicals appear, click on Send to Batch Search. (OK you could have put your own list of chemicals into the Batch search ... another option.)

The screenshot shows the 'Batch Search' page of the EPA CompTox Chemicals Dashboard. The page has a blue header with navigation links: Home, Search, Lists, About, Tools, and a Submit Comments button. Below the header is a welcome message: 'Welcome to the new EPA CompTox Chemicals Dashboard' and a note: 'The new Dashboard is a complete rebuild and is replacing the CompTox Chemicals Dashboard released on July 12th 2020.' The main section is titled 'Batch Search' and is divided into two columns. The left column, labeled '1 Select Input Type(s)', contains a list of input types with checkboxes: Substance Identifiers (checked), Chemical Name, CASRN, InChIKey, DSSTox Substance ID, DSSTox Compound ID, InChIKey Skeleton, MS-Ready Formula(e), Exact Formula(e), and Monoisotopic Mass. The right column, labeled '2 Enter Identifiers to Search', contains a text area with a list of identifiers: DTXSID90880095, DTXSID10880092, DTXSID00880086, DTXSID60214520, DTXSID3031654, DTXSID90880015, DTXSID80880105, and DTXSID60880082. Below the text area are two buttons: '3 DISPLAY ALL CHEMICALS' and 'CHOOSE EXPORT OPTIONS'.

Figure J12. Send a list of chemicals to Batch Search.

Click on the following: Choose Export Options then Choose Export Format then Excel then Abstract Sifter Input file under Enhanced Data Sheets. Then Download. An Excel file will be built. Open it and go to the sheet Abstract Sifter Sheet. Copy the chemicals you want and paste them into the Abstract Sifter either on the Landscape sheet or the CuratedChemicals sheet.

Download Chemical Data, then Download as Excel, then Abstract Sifter Input File (Beta), then (finally) the Download bar. This action will download the chemicals to Excel. Open that file. It will have 2 sheets. Open the one that is called Abstract Sifter. It looks like Figure J15. On the Abstract Sifter Landscape sheet, unhide column A. This is done by clicking on the left border of Column B, then right-clicking to see the menu where you can click on *Unhide*. Paste rows from the downloaded spreadsheet onto the Landscape sheet as in Figure J16.

	A	B	C
1	DSSTOX LINK	PREFERRED NAME	CHEMICAL/ENTITY QUERY
2	DTXSID2031083	Cylindrospermopsin	143545-90-8 OR Cylindrospermopsin
3	DTXSID3031654	Microcystin LR	101043-37-2 OR Microcystin LR OR cyanoginosin LR
4	DTXSID3031656	Microcystin LA	96180-79-9 OR Microcystin LA OR cyanoginosin-LA
5	DTXSID9040974	Azaspiracid	214899-21-5 OR Azaspiracid
6	DTXSID3074313	Saxitoxin	35523-89-8 OR Saxitoxin
7	DTXSID60166611	beta-N-Methylamino-L-alanine	15920-93-1 OR beta-N-Methylamino-L-alanine
8	DTXSID70207660	Decarbamylsaxitoxin	58911-04-9 OR Decarbamylsaxitoxin
9	DTXSID60214520	Gonyautoxin V	64296-25-9 OR Gonyautoxin V
10	DTXSID20274180	L-Domoic acid	14277-97-5 OR L-Domoic acid OR domoic acid
11	DTXSID90423027	palytoxin	77734-91-9 OR palytoxin
12	DTXSID50867064	Anatoxin a	64285-06-9 OR Anatoxin a OR anatoxin I
13	DTXSID60879996	Brevetoxin A	98112-41-5 OR Brevetoxin A
14	DTXSID20879997	Brevetoxin 2	79580-28-2 OR Brevetoxin 2
15	DTXSID40879999	Brevetoxin C	98225-48-0 OR Brevetoxin C
16	DTXSID40880000	Ciguatoxin 1	11050-21-8 OR Ciguatoxin 1
17	DTXSID00880001	Dinophysistoxin 1	81720-10-7 OR Dinophysistoxin 1
18	DTXSID60880002	Okadaic acid	78111-17-8 OR Okadaic acid
19	DTXSID10880012	Maitotoxin	59392-53-9 OR Maitotoxin
20	DTXSID90880015	Lyngbyatoxin-a	70497-14-2 OR Lyngbyatoxin-a
21	DTXSID10880017	Euglenophycin	1219817-69-2 OR Euglenophycin
22	DTXSID60880022	Nodularin	118399-22-7 OR Nodularin
23	DTXSID20880023	Yessotoxin	112514-54-2 OR Yessotoxin
24	DTXSID80880024	Azaspiracid 4	344422-49-7 OR Azaspiracid 4
25	DTXSID60880082	Aplysiatoxin	52659-57-1 OR Aplysiatoxin

Figure J15. Excel view of downloaded chemicals on Abstract Sifter sheet.

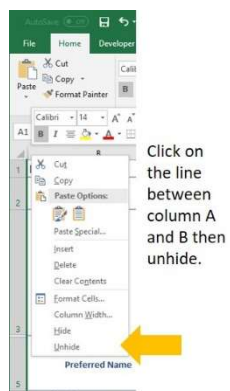


Figure J16. Unhide column A on the Landscape sheet in order to paste the DSSTox number there.

	A	B	C	E	F	G	H
1	Abstract Sifter	Landscape View					
2	Version 7		Update Article Counts	More stuff	Heat Map by column	Heat Map by row	
3				effects OR DNA Damage OR chromosome	s or cancer OR carcinogen* OR precancerous	ion AND (toxicity OR abnormal OR adverse	city OR (Nervous system diseases and chemical
			Subject queries:				
5	(optional) DSSTOX link to Dashboard	Preferred Name	Chemical / Entity query	Genetox	Cancer	ReproTox	NeuroTox
6	DTXSID4020533	1,4-Dioxane	1,4-Dioxane OR 123-91-1				
7	DTXSID90880095	Gonyautoxin 2	60508-89-6 OR Gonyautoxin 2				
8	DTXSID10880092	Pectenotoxin-1	97564-90-4 OR Pectenotoxin-1				
9	DTXSID00880086	Microcystin YR	101064-48-6 OR Microcystin YR				
10	DTXSID60214520	Gonyautoxin 5	64296-25-9 OR Gonyautoxin 5				
11	DTXSID3031654	Microcystin LR	101043-37-2 OR Microcystin LR				
12	DTXSID90880015	Lyngbyatoxin A	70497-14-2 OR Lyngbyatoxin A				
13	DTXSID80880105	Ciguatoxin 3C	148471-85-6 OR Ciguatoxin 3C				
14	DTXSID60880082	Aplysiatoxin	52659-57-1 OR Aplysiatoxin				
15	DTXSID20880023	Yessotoxin	112514-54-2 OR Yessotoxin				
16	DTXSID50880112	Ostreocin-D	163648-25-7 OR Ostreocin-D				
17	DTXSID50867064	Anatoxin a	64285-06-9 OR Anatoxin a				
18	DTXSID10880113	Prymnesin-2	168010-52-4 OR Prymnesin-2				
19	DTXSID60880103	Gymnodimine A	173792-58-0 OR Gymnodimine A				
20	DTXSID40880106	CTX 4A	66231-73-0 OR CTX 4A				
21	DTXSID70880114	Karlotoxin-2	1138665-35-6 OR Karlotoxin-2				
22	DTXSID00880102	Pinnatoin G	NOCAS_880102 OR Pinnatoin G				
23	DTXSID60879996	Brevetoxin A	98112-41-5 OR Brevetoxin A				
24	DTXSID10880012	Maitotoxin	59392-53-9 OR Maitotoxin				
25	DTXSID20880083	19-Bromoaplysiatoxin	66648-18-8 OR 19-Bromoaplysiatoxin				
26	DTXSID20274180	L-Domoic acid	14277-97-5 OR L-Domoic acid				
27	DTXSID60880108	Gambieric acid A	138434-64-7 OR Gambieric acid A				

Figure J17. This is what the sheet will look like after unhiding Column A and pasting the chemicals downloaded from the Dashboard.

Now, enter subject matter queries, or, if you already have queries in place, select the intersecting cells and click on *Update Article Counts*. Click on one of the Heat Map buttons to make it pretty.

	A	B	C	E	F	G	H
1	Abstract Sifter	Landscape View					
2	Version 7		Update Article Counts	More stuff	Heat Map by column	Heat Map by row	
3				effects OR DNA Damage OR chromosome	s or cancer OR carcinogen* OR precancerous	ion AND (toxicity OR abnormal OR adverse	city OR (Nervous system diseases and chemical
			Subject queries:				
5	(optional) DSSTOX link to Dashboard	Preferred Name	Chemical / Entity query	Genetox	Cancer	ReproTox	NeuroTox
6	DTXSID4020533	1,4-Dioxane	1,4-Dioxane OR 123-91-1	64	171	5	127
7	DTXSID90880095	Gonyautoxin 2	60508-89-6 OR Gonyautoxin 2	2	9	0	20
8	DTXSID10880092	Pectenotoxin-1	97564-90-4 OR Pectenotoxin-1	0	16	0	1
9	DTXSID00880086	Microcystin YR	101064-48-6 OR Microcystin YR	81	1542	76	87
10	DTXSID60214520	Gonyautoxin 5	64296-25-9 OR Gonyautoxin 5	0	0	0	6
11	DTXSID3031654	Microcystin LR	101043-37-2 OR Microcystin LR	107	1884	89	109
12	DTXSID90880015	Lyngbyatoxin A	70497-14-2 OR Lyngbyatoxin A	1	41	0	0
13	DTXSID80880105	Ciguatoxin 3C	148471-85-6 OR Ciguatoxin 3C	0	0	0	7
14	DTXSID60880082	Aplysiatoxin	52659-57-1 OR Aplysiatoxin	4	95	0	3
15	DTXSID20880023	Yessotoxin	112514-54-2 OR Yessotoxin	10	75	0	9
16	DTXSID50880112	Ostreocin-D	163648-25-7 OR Ostreocin-D	0	1	0	0
17	DTXSID50867064	Anatoxin a	64285-06-9 OR Anatoxin a	10	216	6	64
18	DTXSID10880113	Prymnesin-2	168010-52-4 OR Prymnesin-2	0	0	0	0
19	DTXSID60880103	Gymnodimine A	173792-58-0 OR Gymnodimine A	0	12	0	6
20	DTXSID40880106	CTX 4A	66231-73-0 OR CTX 4A	2	5	0	3
21	DTXSID70880114	Karlotoxin-2	1138665-35-6 OR Karlotoxin-2	0	1	0	1
22	DTXSID00880102	Pinnatoin G	NOCAS_880102 OR Pinnatoin G	1	2	0	4
23	DTXSID60879996	Brevetoxin A	98112-41-5 OR Brevetoxin A	1	3	0	1
24	DTXSID10880012	Maitotoxin	59392-53-9 OR Maitotoxin	2	74	1	73
25	DTXSID20880083	19-Bromoaplysiatoxin	66648-18-8 OR 19-Bromoaplysiatoxin	0	0	0	0
26	DTXSID20274180	L-Domoic acid	14277-97-5 OR L-Domoic acid	87420	359844	14592	169354
27	DTXSID60880108	Gambieric acid A	138434-64-7 OR Gambieric acid A	0	3	0	1
28	DTXSID90880111	Ovatoin A	1009813-91-5 OR Ovatoin A	0	2	0	0
29	DTXSID80880100	Decarbamoylneoxaitoxin	66883-58-9 OR Decarbamoylneoxaitoxin	0	0	0	3

Figure J18. Downloaded chemicals and queries with subject matter queries.

Now you have an overview of your chemicals and what literature is out in PubMed for them. Take advantage of the iterative nature of the Abstract Sifter to query, sift, read, note as much as you need.

M. In-depth discussions

1. MeSH Mining in-depth

Most PubMed citations are enhanced by the addition of MeSH indexing terms. This indexing is performed by a combination of automated text-mining and manual curation. One of the entities indexed is substance name. Substances (like chemicals) discussed in the abstract will likely have the chemical name annotated, often along with a family name for that chemical. Note: publications that discuss lots

of chemicals in the text (i.e. “we studied 25 pesticides …”) will not have every chemical annotated but will likely have the term Pesticides in the annotations.

Often, too, annotations will have subheadings of interest. In the realm of chemicals, these subheadings can indicate whether the toxicity (adverse effects, poisoning) of the chemical is being discussed or the therapeutic use of the chemical.

The MeSH Mine sheet pulls out substances from the MeSH terms, finds their associated entities and writes them with counts and flags to the MeSHMine sheet.

Here we’ll look at how the MeSH mining function works. To start, let’s go to PubMed and find the article with PMID 28178360. Here’s a link: [Melatonin protects oocyte quality from Bisphenol A-induced deterioration in the mouse - PubMed \(nih.gov\)](#)

Read the abstract and then on the right, click on MeSH Terms. You’ll see the results to the right. (Figure M-1)

You’ll see a variety of MeSH terms at the top of the list: substances, species, organs, processes, anatomy, cells, etc.

The substance list at the bottom is what the MeSH mine function focuses on. The Abstract Sifter extracts those terms to put on the MeSHMine sheet.

Figuring out whether the substance is discussed therapeutically or for its adverse effects is slightly more complicated. The Abstract Sifter code looks for subheadings or qualifiers associated with a substance (or its parent family

term).

MeSH terms
> Animals
> Benzhydryl Compounds / toxicity*
> Female
> Fertilization / drug effects*
> Male
> Meiosis / drug effects*
> Melatonin / pharmacology*
> Metalloproteases / metabolism
> Mice
> Mice, Inbred ICR
> Oocytes / metabolism*
> Oocytes / pathology
> Phenols / toxicity*
> Receptors, Cell Surface / metabolism
> Sperm-Ovum Interactions / drug effects*
> Spindle Apparatus / metabolism
> Spindle Apparatus / pathology
> Zona Pellucida / metabolism
> Zona Pellucida / pathology
> Zona Pellucida Glycoproteins / metabolism
Substances
> Benzhydryl Compounds
> Phenols
> Receptors, Cell Surface
> Zona Pellucida Glycoproteins
> Zp2 protein, mouse
> folate receptor 4, mouse
> Astl protein, mouse
> Metalloproteases
> Melatonin
> bisphenol A

Figure M-1. MeSH term and substance annotations.

2. Run in Batch in-depth

In this section we'll walk through the batch processing capabilities of the Abstract Sifter starting with the Landscape sheet.

The batch processing features allow the user to build a corpus on the Main sheet from more than one query, appending the results of one query to another. You can do this manually, simply by running a query, then running another query and selecting the append option. But let's say you have many queries and you want a more automated approach to building a corpus. We'll walk through an example here that puts together all the pieces.

So, let's start with the Landscape sheet. In this example we're looking at a set of toxins and queries about toxicity types. (Figure M-2) We decide to get all the literature on neurotoxicity for this set of chemicals. We could double-click on one cell, run the query and send the results to Main, then double-click on the next cell and send to Main with the Append option, etc. A better approach is to use the batch run capabilities. First, select the cells you want to get results for.

	A	B	C	E	F	G	H
1	Abstract Sifter	Landscape View					
2	Version 7		Update Article Counts More stuff Heat Map by column Heat Map by row				
3			Subject queries:	effects OR DNA Damage OR chromosome	s or cancer OR carcinogen* OR precancerous	ion AND (toxicity OR abnormal OR adverse	ty OR (Nervous system diseases and chemically
5	(optional) DSSTOX link to Dashboard	Preferred Name	Chemical / Entity query	Genetox	Cancer	ReproTox	NeuroTox
6	DTXSID90880095	Gonyautoxin 2	60508-89-6 OR Gonyautoxin 2	2	9	0	20
7	DTXSID10880092	Pectenotoxin-1	97564-90-4 OR Pectenotoxin-1	0	16	0	1
8	DTXSID00880086	Microcystin YR	101064-48-6 OR Microcystin YR	81	1542	76	87
9	DTXSID60214520	Gonyautoxin 5	64296-25-9 OR Gonyautoxin 5	0	0	0	6
10	DTXSID3031654	Microcystin LR	101043-37-2 OR Microcystin LR	107	1884	89	109
11	DTXSID00880102	Pinnatoin G	NOCAS_880102 OR Pinnatoin G	1	2	0	4
12	DTXSID60879996	Brevetoxin A	98112-41-5 OR Brevetoxin A	1	3	0	1
13	DTXSID10880012	Maitotoxin	59392-53-9 OR Maitotoxin	2	74	1	73
14	DTXSID20880083	19-Bromoaplysiatoxin	66648-18-8 OR 19-Bromoaplysiatoxin	0	0	0	0
15	DTXSID20274180	L-Domoic acid	14277-97-5 OR Domoic acid	15	72	34	361
16	DTXSID60880108	Gambieric acid A	138434-64-7 OR Gambieric acid A	0	3	0	1
17	DTXSID90880111	Ovatoin A	1009813-91-5 OR Ovatoin A	0	2	0	0
18	DTXSID80880100	Decarbamoylneosaxitoxin	68683-58-9 OR Decarbamoylneosaxitoxin	0	0	0	3

Figure M-2. Landscape sheet with counts.

	A	B	C	E	F	G	H
1	Abstract Sifter	Landscape View					
2	Version 7		Update Article Counts	More stuff	Heat Map by column	Heat Map by row	
3							
4							
5	(optional) DSSTOX link to Dashboard	Preferred Name	Chemical / Entity query	Genetox	Cancer	ReproTox	NeuroTox
6	DTXSID90880095	Gonyautoxin 2	60508-89-6 OR Gonyautoxin 2	2	9		20
7	DTXSID10880092	Pectenotoxin-1	97564-90-4 OR Pectenotoxin-1		16		1
8	DTXSID00880086	Microcystin YR	101064-48-6 OR Microcystin YR	81	1542	76	87
9	DTXSID60214520	Gonyautoxin 5	64296-25-9 OR Gonyautoxin 5				6
10	DTXSID3031654	Microcystin LR	101043-37-2 OR Microcystin LR	107	1884	89	109
11	DTXSID00880102	Pinnatoin G	NOCAS_880102 OR Pinnatoin G	1	2		4
12	DTXSID60879996	Brevetoxin A	98112-41-5 OR Brevetoxin A	1	3		1
13	DTXSID10880012	Maitotoxin	59392-53-9 OR Maitotoxin	2	74	1	73
14	DTXSID20880083	19-Bromoaplysiatoxin	66648-18-8 OR 19-Bromoaplysiatoxin				
15	DTXSID20274180	L-Domoic acid	14277-97-5 OR Domoic acid	15	72	34	361
16	DTXSID60880108	Gambieric acid A	138434-64-7 OR Gambieric acid A		3		1
17	DTXSID90880111	Ovatoxin A	1009813-91-5 OR Ovatoxin A		2		
18	DTXSID80880100	Decarbamoylneosaxitoxin	68683-58-9 OR Decarbamoylneosaxitoxin				3
19							
20							

Select cells you want results for on Main sheet

Figure M-3. Select the cells to be sent to the Log sheet for batch running.

Next, click on the More Stuff button above and follow that by clicking on Send to Log.

Then click on More stuff then Send to Log

Figure M-4. More stuff and Send to Log functions.

That action sends queries to the top of the Log sheet. Here's what they look like:

A		B	C	F
1	Log / Batch		← Main Run in batch	Note: Feel free to delete
2	Date	Record Ct	Query Used (double-click on query to rerun)	Batch Tag
3	7/13/2022 13:36	3	(68683-58-9 OR Decarbamoylneosaxitoxin) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((n	Decarbamoylneosaxitoxin
4	7/13/2022 13:36	0	(1009813-91-5 OR Ovatoxin A) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR b	Ovatoxin A
5	7/13/2022 13:36	1	(138434-64-7 OR Gambieric acid A) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons C	Gambieric acid A
6	7/13/2022 13:36	361	(14277-97-5 OR Domoic acid) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR br	Domoic acid
7	7/13/2022 13:36	0	(66648-18-8 OR 19-Bromoaplysiatoxin) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurc	19-Bromoaplysiatoxin
8	7/13/2022 13:36	73	(59392-53-9 OR Maitotoxin) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR brai	Maitotoxin
9	7/13/2022 13:36	1	(98112-41-5 OR Brevetoxin A) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR br	Brevetoxin A
10	7/13/2022 13:36	4	(NOCAS_880102 OR Pinnatoxin G) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons C	Pinnatoxin G
11	7/13/2022 13:36	110	(101043-37-2 OR Microcystin LR) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR	Microcystin LR
12	7/13/2022 13:36	6	(64296-25-9 OR Gonyautoxin 5) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR	Gonyautoxin 5
13	7/13/2022 13:36	87	(101064-48-6 OR Microcystin YR) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR	Microcystin YR
14	7/13/2022 13:36	1	(97564-90-4 OR Pectenotoxin-1) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR	Pectenotoxin-1
15	7/13/2022 13:36	20	(60508-89-6 OR Gonyautoxin 2) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR	Gonyautoxin 2

Figure L2.4 Log sheet with queries sent from the Landscape sheet.

There are a couple of things to notice here. First, the queries are in italics. That's just to indicate that they came from the Landscape sheet. They also have batch tags. The batch tag by default is the value on column B of the Landscape sheet, in this case the preferred name of the chemical. Feel free to change it.

To run each of the queries and put all the results together on the Main sheet, select rows and click on Run in Batch. A form appears asking whether you want to clear results from Main before the batch run or if you want to append to those results. (Figure M-5)

A		B	C	F
1	Log / Batch		← Main Run in batch	Note: Feel free to delete
2	Date	Record Ct	Query Used (double-click on query to rerun)	Batch Tag
3	7/13/2022 13:36	3	(68683-58-9 OR Decarbamoylneosaxitoxin) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((n	Decarbamoylneosaxitoxin
4	7/13/2022 13:36	0	(1009813-91-5 OR Ovatoxin A) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR b	Ovatoxin A
5	7/13/2022 13:36	1	(138434-64-7 OR Gambieric acid A) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons C	Gambieric acid A
6	7/13/2022 13:36	361	(14277-97-5 OR Domoic acid) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR br	Domoic acid
7	7/13/2022 13:36	0	(66648-18-8 OR 19-Bromoaplysiatoxin) AND ((neurotoxicity C	19-Bromoaplysiatoxin
8	7/13/2022 13:36	73	(59392-53-9 OR Maitotoxin) AND ((neurotoxicity OR (Nervau	Maitotoxin
9	7/13/2022 13:36	1	(98112-41-5 OR Brevetoxin A) AND ((neurotoxicity OR (Nervo	Brevetoxin A
10	7/13/2022 13:36	4	(NOCAS_880102 OR Pinnatoxin G) AND ((neurotoxicity OR (N	Pinnatoxin G
11	7/13/2022 13:36	110	(101043-37-2 OR Microcystin LR) AND ((neurotoxicity OR (Ne	Microcystin LR
12	7/13/2022 13:36	6	(64296-25-9 OR Gonyautoxin 5) AND ((neurotoxicity OR (Ner	Gonyautoxin 5
13	7/13/2022 13:36	87	(101064-48-6 OR Microcystin YR) AND ((neurotoxicity OR (Ne	Microcystin YR
14	7/13/2022 13:36	1	(97564-90-4 OR Pectenotoxin-1) AND ((neurotoxicity OR (Ne	Pectenotoxin-1
15	7/13/2022 13:36	20	(60508-89-6 OR Gonyautoxin 2) AND ((neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR	Gonyautoxin 2
16	5/23/2022 16:43	2460	(Anthozoa OR coral) AND (chemical toxicity OR drug effects)	

Figure M-5. To run all queries and combine results - select queries then click no Run in batch.

When the batch run is complete, you'll have all the records from each query on the Main sheet. Now let's look at the Main sheet and see how we know which records are from which query. Take a look at column R on the Main sheet. It has the batch tags.

There's another feature that's helpful here. On the Main sheet, click on More things, then Summarize. A new sheet with the results summary by batch is inserted. (Figure K2.6)

R	Example of Batch tags and Main sheet summary by Batch tag	
	A	B
	1	Summary of Main Sheet
	2	
Batch Tag	3	Batch Tag Count of Doc ID
Decarbamoylneosaxi	4	Brevetoxin A 1
Decarbamoylneosaxi	5	Decarbamoylneosaxitoxin 3
Decarbamoylneosaxi	6	Gambieric acid A 1
Gambieric acid A	7	Gonyautoxin 2 20
L-Domoic acid	8	Gonyautoxin 5 6
L-Domoic acid	9	L-Domoic acid 361
L-Domoic acid	10	Maitotoxin 73
L-Domoic acid	11	Microcystin LR 110
L-Domoic acid	12	Microcystin YR 87
L-Domoic acid	13	Pectenotoxin-1 1
L-Domoic acid	14	Pinnatoin G 4
L-Domoic acid	15	Grand Total 667

Figure M-6. Counting and summarizing rows by batch tags.

Many times a set of queries could produce duplicate publications on the Main sheet. To eliminate duplicates, click on More things button, then the Dups button. Note that the counts of publications by batch tag will change and there is no control over which record (with which batch tag) is deleted by the automatic method.

In summary, running in batch is a great way to build a large corpus or a complex corpus. Note, too, that although this discussion showed running queries sent from Landscape in batch, you can run any queries on the Log sheet in batch mode.

N. Troubleshooting

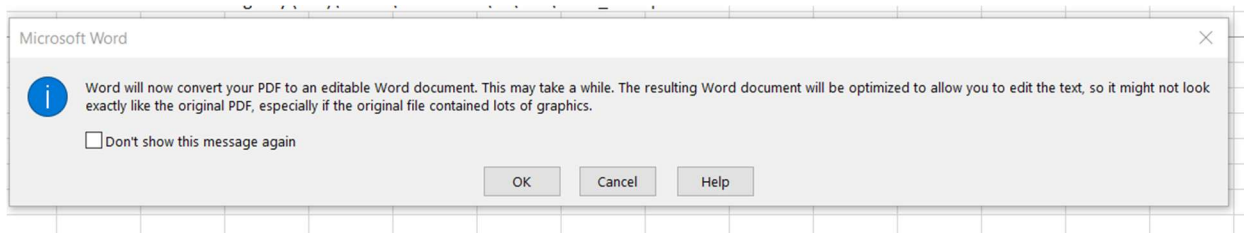
1. Extracting from pdfs

The methodologies used to extract from pdfs take advantage of file types. The table extraction feature, for instance, opens the pdf in Microsoft Word. Word automatically transforms the pdf into a Word document (performs some kind of OCR if necessary) and will try to recognize tables and turn them into table objects. The Abstract Sifter locates each of the table objects and iterates through the columns and rows, copies the data, and writes it to the Sifter.

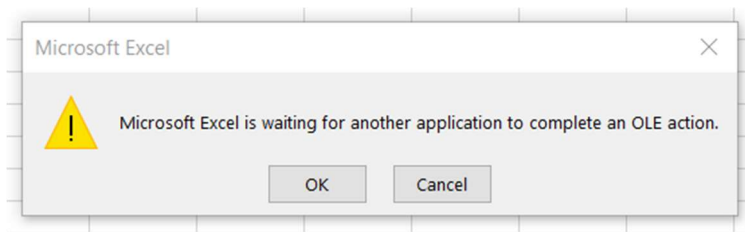
Sometimes something happens and the pdf gets opened in Word and hits an error and never gets closed. This can mean that pdf is not accessible, and you start to accumulate Word processes. So if you hit a snag, do a `ctl-alt-delete`, click on Task Manager, then look for any running Word processes and cancel them.

The feature that extracts figures works differently from the table extraction feature. It actually creates a new version of the pdf as a zip file, then opens the zip directory and copies the figures. This works in the EPA environment. It may not work in yours. Please let me know.

(Note – you may be notified that Word is opening your pdf file and is that OK. Answer yes. And turn off notifications.)



Another note – if your pdf is big, you might get this message:



Keep clicking on OK. Eventually you should get the count of tables:

O. Technical appendix

PubChem options, subroutines, and API calls



- 1 Get DTXSID from PubChem using chemical name
- 2 Get DTXSID from PubChem using SMILES
- 3 Get PubChem name using SMILES
- 4 Get PubChem CID using SMILES
- 5 Get PubChem CID using name or CAS or synonym
- 6 Get DTXSID from PubChem using CID
- 7 Get EC number (ECHA) from PubChem using CID
- 8 Get MeSH Name from PubChem using CID
- 9 Get CASRN from PubChem using CID
- 10 Get PubChem patent counts using CID
- 11 Get PubChem PubMed article counts using CID
- 12 Get PubChem vendor counts using CID
- 13 PubChem - is Sigma a vendor?

#	Function / API calls / Notes
1	getDTXSIDfromPubChemviasyn https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/name/Imidaclothiz/synonyms/XML
2	getDTXSIDviaSMILES https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/smiles/CCOP(=O)(OCC)OCC/cids/xml https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/6535/synonyms/XML
3	getPubChemNameviaSMILES getpcchemname(smiles) https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/smiles/CCOP(=O)(OCC)OCC/cids/xml https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/6535/synonyms/XML
4	getPubChemCIDviaSMILES(colnum) https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/smiles/CCOP(=O)(OCC)OCC/cids/xml

5	<p>getCIDfromPubChemviasyn(colnum)</p> <p>getPubChemIDsinSynlist</p> <p>https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/name/cyclopamine/synonyms/XML</p> <p>Note: cid is top of list</p>
6	<p>getDSSToxIDviaCID(colnum)</p> <p><a "="" &="" cid="" href="https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/" synonyms="" xml"="">https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/" & cid & "/synonyms/XML</p>
7	<p>getECfromPubChemviaCID(colnum)</p> <p>genPugViewCall(cid, "EC")</p> <p>https://pubchem.ncbi.nlm.nih.gov/rest/pug_view/data/compound/3121/XML/</p> <p>Note: look for "European Community (EC) Number" section</p>
8	<p>ccdgetSomething(colnum, "pubchemMeSHName")</p> <p>getPubChemMeSHName(myquery)</p> <p>https://pubchem.ncbi.nlm.nih.gov/rest/pug_view/literature/compound/338/json/</p> <p>https://www.ncbi.nlm.nih.gov/mesh?term="Salicylic Acid"[mh]&report=xml</p> <p>https://www.ncbi.nlm.nih.gov/mesh?term="bisphenol a"[nm]&report=xml</p>
9	<p>getCASfromPubChemviaCID(colnum)</p> <p>genPugViewCall</p> <p>https://pubchem.ncbi.nlm.nih.gov/rest/pug_view/data/compound/338/XML/</p> <p>Look for <TOCHeading>CAS</TOCHeading></p>
10	<p>ccdgetSomething(colnum, "pubchempatentct")</p> <p>getPubChempatentCts(myquery)</p> <p>https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/338/xrefs/PatentID/XML</p> <p>https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/6623/xrefs/PatentID/TXT</p> <p>The patent node count is retrieved. This is slow to retrieve and load. A time out error happens when patent count is approx. > 100000</p>
11	<p>ccdgetSomething(colnum, "pubchemartct")</p>

	https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/74483/xrefs/PubMedID/XML
12	<p>ccdgetSomething(colnum, "pubchemvendorct")</p> <p>https://pubchem.ncbi.nlm.nih.gov/rest/pug_view/categories/compound/6623/xml</p> <p>Look for Category>Chemical Vendors<</p> <p>Each vendor listed is checked against a list of valid vendors:</p> <p>https://pubchem.ncbi.nlm.nih.gov/rest/pug/sourcetable/all/csv/?response_type=display</p> <p>Those designated as legacy are omitted.</p>
13	<p>ccdgetSomething(colnum, "pubchemSigmaisVendor")</p> <p>getPubChemvendorCts(myquery, "sigmaisvendor")</p> <p>Note: checks for "sigma" in vendor list</p>

P. Bibliography

1. Kim, S., P. A. Thiessen, T. Cheng, B. Yu, B. A. Shoemaker, J. Wang, E. E. Bolton, Y. Wang, and S. H. Bryant. "Literature Information in Pubchem: Associations between Pubchem Records and Scientific Articles." *J Cheminform* 8 (2016): 32; DOI: 10.1186/s13321-016-0142-6.
2. Wei, C. H., A. Allot, R. Leaman, and Z. Lu. "Pubtator Central: Automated Concept Annotation for Biomedical Full Text Articles." *Nucleic Acids Res* 47, no. W1 (2019): W587-W93; DOI: 10.1093/nar/gkz389.
3. Wei, C. H., A. Allot, P. T. Lai, R. Leaman, S. Tian, L. Luo, Q. Jin, Z. Wang, Q. Chen, and Z. Lu. "Pubtator 3.0: An Ai-Powered Literature Resource for Unlocking Biomedical Knowledge." *ArXiv* (2024).
4. Kim, S., J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton. "Pubchem 2023 Update." *Nucleic Acids Res* 51, no. D1 (2023): D1373-D80; DOI: 10.1093/nar/gkac956.
5. Dix, D. J., K. A. Houck, M. T. Martin, A. M. Richard, R. W. Setzer, and R. J. Kavlock. "The Toxcast Program for Prioritizing Toxicity Testing of Environmental Chemicals." *Toxicol Sci* 95, no. 1 (2007): 5-12; DOI: 10.1093/toxsci/kfl103.
6. Williams, A. J., C. M. Grulke, J. Edwards, A. D. McEachran, K. Mansouri, N. C. Baker, G. Patlewicz, I. Shah, J. F. Wambaugh, R. S. Judson, and A. M. Richard. "The Comptox Chemistry Dashboard: A Community Data Resource for Environmental Chemistry." *J Cheminform* 9, no. 1 (2017): 61; DOI: 10.1186/s13321-017-0247-6.
7. Yang, C., A. Tarkhov, J. Maruszyk, B. Bienfait, J. Gasteiger, T. Kleinoeder, T. Magdziarz, O. Sacher, C. H. Schwab, J. Schwoebel, L. Terflath, K. Arvidson, A. Richard, A. Worth, and J. Rathman. "New Publicly Available Chemical Query Language, Csrml, to Support Chemotype Representations for Application to Data Mining and Modeling." *J Chem Inf Model* 55, no. 3 (2015): 510-28; DOI: 10.1021/ci500667v.
8. Gamble, J. T., K. Hopperstad, and C. Deisenroth. "The Devtox Germ Layer Reporter Platform: An Assay Adaptation of the Human Pluripotent Stem Cell Test." *Toxics* 10, no. 7 (2022); DOI: 10.3390/toxics10070392.

Log

02/10/2025 Added text about code signing.