

## **Addition of Average MW to Web-Based Version of SciFinder**

*July 28, 2012, J Little, Eastman Chemical Company*

They have recently added the ability to search by **average MW** to the web-based version of SciFinder. This is a very nice addition since one no longer has to employ the somewhat complicated STN Express interface.

There are a few restrictions. The maximum number of answers a "Reference," "Substance," or "Reaction" answer set can be sorted is 20,000. Although one cannot sort an answer set larger than this limit, one can still "Refine" by key word and answer set >20,000. There is a "work-around" to break an answer set that is greater than 20,000 into smaller sizes that can then be sorted by the number of references. That option is discussed at the end of this note.

Here is an example that I had used in my initial ASMS article to demonstrate the new capability. Previously, one could only search by average MW in the STN Express interface. Hopefully at some time in the future they will include the capability to search by monoisotopic mass. However, in the meantime, this capability will still be very useful!

All average molecular weights in the CAS Registry are only listed to two significant decimals to the right of the decimal. On our current older instruments, I usually use a 70 ppm window for average molecular weight. The average molecular weight must be calculated manually by considering the m/z values and intensities of all major isotopes for the molecular ion or ion adduct for the component of interest.

In the article, page 9, I had shown an example for an unknown compound with average MW of 441.60:

[http://www.littledomain.com/james/files/known\\_unknowns\\_article.pdf](http://www.littledomain.com/james/files/known_unknowns_article.pdf)

SciFinder®

Explore References   Explore Substances   Explore Reactions

Welcome Jim Little | Sign Out

## Explore Substances

Chemical Structure  
Markush  
Molecular Formula  
**Property** NEW  
Substance Identifier

**Property**

Select the category and enter an appropriate value or range.

Experimental

Select Property...

*Examples: Individual value as 44, range as 25-35, or open ended range as >125 or <12*

Predicted

Select Property...

*Examples: Individual value as 44, range as 25-35, or open ended range as >125 or <12*

- Select Property...
- Bioconcentration Factor
- Boiling Point (°C)
- Density (g/cm<sup>3</sup>)
- Enthalpy of Vaporization (kJ/mol)
- Flash Point (°C)
- Freely Rotatable Bonds
- H Donor/Acceptor sum
- H Acceptors
- H Donors
- Koc
- logD
- logP
- Mass Intrinsic Solubility (g/L)
- Mass Solubility (g/L)
- Molar Intrinsic Solubility (mol/L)
- Molar Solubility (mol/L)
- Molar Volume (cm<sup>3</sup>/mol)
- Molecular Weight**
- pKa
- Polar Surface Area (Å<sup>2</sup>)
- Vapor Pressure (Torr)

Search the Average MW 441.60 (actually enter window of 441.57-441.63):

SciFinder®

Welcome Jim Little | Sign Out

Explore References   Explore Substances   Explore Reactions

Add KMP Alert   Property "Predicted - Molecular Weight, ..." > substances (15172)

### Explore Substances

Chemical Structure  
Markush  
Molecular Formula  
**Property** NEW  
Substance Identifier

**Property** ⚙

Select the category and enter an appropriate value or range.

Experimental

Select Property...   Value or Range ⚙

*Examples: Individual value as 44, range as 25-35, or open ended range as >125 or <125*

Predicted

Molecular Weight   Value or Range ⚙

441.57-441.63

*Examples: Individual value as 44, range as 25-35, or open ended range as >125 or <125*

One will get the following display *after* sorting by:

SciFinder®

Welcome Jim Little | Sign Out

Explore References   Explore Substances

Add KMP Alert   Property "Predicted - Molecular Weight, ..." > substances (15347)

Substances   Get References   Get Reactions   Tools   Send to SciPlanner

15347 Substances   0 Selected

Select All   Deselect All   Sort by: CAS Registry Number

CAS Registry Number  
Number of References  
Molecular Weight  
Molecular Formula

1. Substance Detail  
1383306-11-3

~0

Me   i-Pr   CC(C)C(=O)S(=O)(=O)C

SciFinder® Explore References Explore Substances Explore Reactions

Welcome Jim Little | Sign Out

Add KMP Alert Property "Predicted - Molecular Weight, ..." > substances (15172)

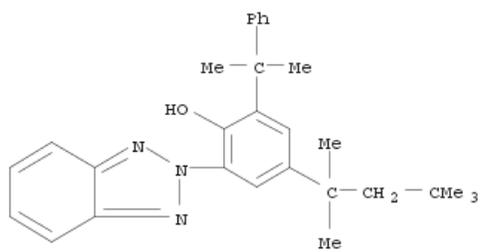
Substances Get References Get Reactions Tools Send to SciPlanner

15172 Substances 0 Selected

Select All Deselect All Sort by: Number of References ↓

1. Substance Detail  
73936-91-1

~146   



**C<sub>29</sub> H<sub>35</sub> N<sub>3</sub> O**  
Phenol, 2-(2H-benzotriazol-2-yl)-6-(1-methyl-1-phenylethyl)-4-(1,1,3,3-tetramethylbutyl)-

The top hit, CAS No. 73936-91-1 was the correct hit!

Another alternative for solving this "known unknown" would have been to get all the references:

Welcome Jim Little | Sign Out

Add KMP Alert Property "Predicted - Molecular Weight, ..." > substances (15347)

Substances Get References Get Reactions Tools

15347 Substances 0 Selected

Select All Deselect All Sort by: CAS Registry Number ↓

1. Substance Detail  
1383306-11-3

Get References

Retrieve references for:

- All substances
- Selected substances

Limit results to:

- Adverse Effect, including toxicity
- Analytical Study
- Prophetics in Patents
- Preparation

One could also have gotten all references for all the compounds (15347) from the above search. This would yield 4406 references. At that point, one can refine by UV stabilizer using the menu on the right side of the screen:

Analysis
Refine

**Refine by:** ⓘ

- Research Topic
- Author Name
- Company Name
- Document Type
- Publication Year
- Language
- Database

---

**Research Topic**

uv stabilizer

Examples:

*The effect of antibiotic residues on dairy products*

*Photocyanation of aromatic compounds*

[Refine](#)

All the top references had the correct compound highlighted. For example, opening the first reference shows the correct compound highlighted:

[Explore References](#)
[Explore Substances](#)
[Explore Reactions](#)

Welcome Jim Little | [Sign Out](#)

[Add KMP Alert](#) [Property "Predicted - Molecular Weight, ..."](#) > [substances \(15172\)](#) > [get references \(4399\)](#) > [refine "uv stabilizer" \(94\)](#) >

References
[Get Substances](#)
[Get Reactions](#)
[Get Related](#)
[Tools](#)
[Send to SciPlanner](#)

94 References    0 Selected

[Select All](#)   [Deselect All](#)   Sort by: Accession Number ↓

---

1. **UV-absorbing coating materials and LED devices and LED lamps therewith** 🔍 [Full Text](#)

By Okamoto, Akira; Hashimoto, Yoshiyuki  
 From Jpn. Kokai Tokkyo Koho (2012), JP 2012111811 A 20120614. | Language: Japanese, Database: CAPLUS

The coating materials contain (A) F-contg. copolymers having fluoroolefin-based units, (B) org. solvents, (C) benzotriazole- and/or hydr emitting surfaces of LED chips, light-outgoing surfaces of phosphors, reflective surfaces of reflectors, and/or light-outgoing surfaces of PET film was coated with a compn. contg. LF 200 (chlorot...

88795-12-4P Chlorotrifluoroethylene-cyclohexyl vinyl ether-ethyl vinyl ether-4-hydroxybutyl vinyl ether copolymer 🔍

LF 200; UV-absorbing coating materials for LED packages and LED lamps

Industrial manufacture; Reactant; Preparation; Reactant or reagent

83044-89-7 Octyl 3-[3-tert-butyl-4-hydroxy-5-(5-chloro-2H-benzotriazol-2-yl)phenyl]propionate mixture 🔍

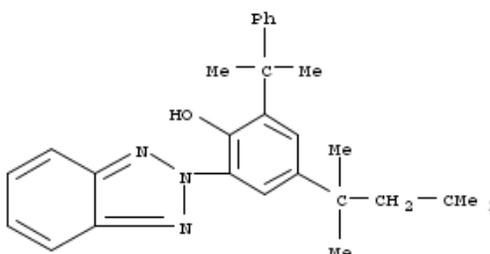
Tinuvin 109, UV absorbers; UV-absorbing coating materials for LED packages and LED lamps

Modifier or additive use; Uses

501-52-0 Benzenepropanoic acid 🔍

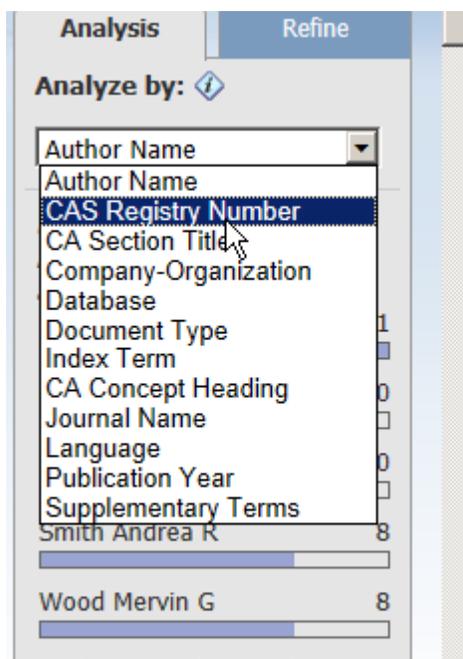
70321-86-7 2-(2H-Benzotriazol-2-yl)-4,6-bis(1-methyl-1-phenylethyl)phenol 🔍

73936-91-1 2-(2H-Benzotriazol-2-yl)-6-(1-methyl-1-phenylethyl)-4-(1,1,3,3-tetramethylbutyl)phenol 🔍

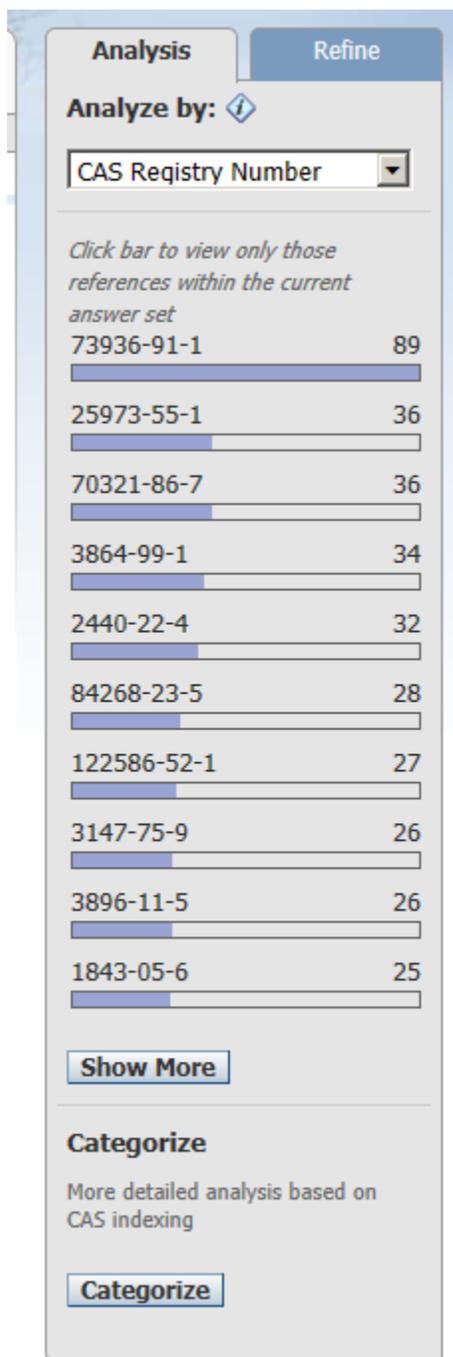


204848-45-3 Tinuvin 479 🔍

At this point instead of manually looking through all the compounds in the top references, one can just select the Analysis window tab with “analyze by” “CAS Registry No.” option:



To get the following display:



Again the top hit with 89 associated references is the correct compound. This latter display will also list other compounds that tend to be used with this UV stabilizer or in place of this stabilizer.

Another way to accomplish the same results is after getting all the references, click on "Categorize" at bottom of the window as shown below:

Analysis Refine

Analyze by: 

CAS Registry Number 

*Click bar to view only those references within the current answer set*

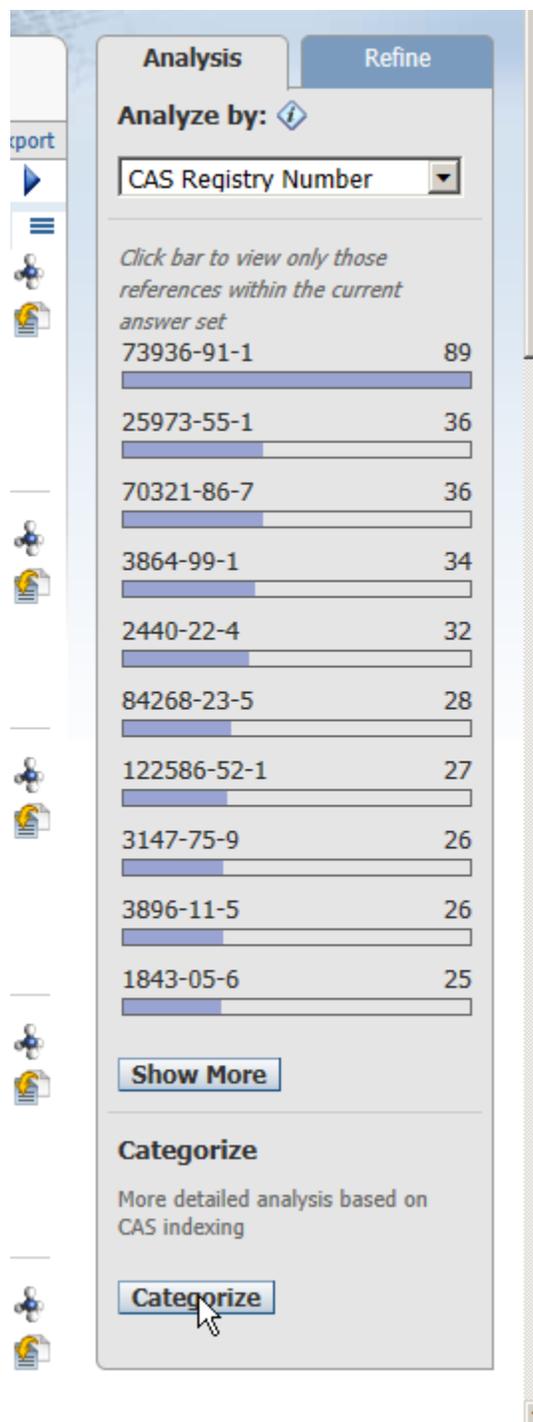
73936-91-1	89
25973-55-1	36
70321-86-7	36
3864-99-1	34
2440-22-4	32
84268-23-5	28
122586-52-1	27
3147-75-9	26
3896-11-5	26
1843-05-6	25

[Show More](#)

**Categorize**

More detailed analysis based on CAS indexing

[Categorize](#)



Then select substances:

**Categorize** ⓘ

1. Select a heading and category.      2. Select index terms of interest.

Category Heading ⓘ	Category ⓘ	Index Terms ⓘ
All	Substances (1852)	1 2 3 4 ... 19 ▶
Technology	Topics (151) ⓘ	Select All Deselect All
General chemistry	Searched substances (7)	<input type="checkbox"/> Tinuvin 928 89 ▲
Polymer chemistry		<input type="checkbox"/> 2-(2-Hydroxy-3,5-di-tert- amylphenyl)-2H- benzotriazole 36
Synthetic chemistry		<input type="checkbox"/> 2-(2-Hydroxy-3,5-di- $\alpha$ - cumylphenyl)-2H- benzotriazole 36
Physical chemistry		<input type="checkbox"/> 5-Chloro-2-(2-hydroxy-3,5- di-tert-butylphenyl)-2H- benzotriazole 34
Biotechnology		<input type="checkbox"/> 2-(2-Hydroxy-5- methylphenyl)-2H- benzotriazole 32
Genetics & protein chemistry		<input type="checkbox"/> Polyesters 31
Catalysis		<input type="checkbox"/> Tinuvin 384 28 ▼
Environmental chemistry		<input type="checkbox"/> Tinuvin 122 27 ▼
Biology		

All > Substances

This again has other compounds found in the same references including stabilizers, the type of polymer, etc.

**Handling Answer Sets >20,000 Entries:** As noted earlier, the maximum number of answers a “Reference,” “Substance,” or “Reaction” answer set can be sorted is 20,000. Although one cannot sort an answer set larger than this limit, one can still “Refine” by key word and answer set >20,000.

There is a “work-around” to break an answer set that is greater than 20,000 into smaller sizes that can then be sorted by the number of references. *This is not trivial, so be patient.* One must save the answer set. I will demonstrate this with the above example in which I opened the mass window to get >20,000 hits by using 441.56-441.64 to search by average molecular weight. As you can see, the option to “Sort by:” is greyed out below:

SciFinder®

Welcome Jim Little | Sign Out

Property "Predicted - Molecular Weight, ..." > substances (22214)

Substances | Get References | Get Reactions | Tools | Send to SciPlanner

22214 Substances 0 Selected

Select All Deselect All | Sort by: CAS Registry Number

1. Substance Detail  
1383306-11-3

Save this work list (right side of window below):

Get References | Get Reactions | Tools | Send to SciPlanner

0 Selected

Save answers to the CAS server. To save answers to your desktop, use Export.

Save | Print | Export

All | Sort by: CAS Registry Number

Answers per Page [15] 1 2 3 4 5 6 ... 1481

I saved the list as test which will save only the first 20,000 entries of the 22,214 initially obtained:

Save This Answer Set ⓘ

Only 20,000 answers can be saved at one time.

\* Required

Save:

- All answers
- Only selected answers

Title: \*

first 20000

Description:

OK Cancel

Then use tools/combine answer sets:

Substances Get References Get Reactions Tools Select to view a items

22214 Substances 0 Selected Sort by: CAS Registry Number

1. Substance Detail  
1383306-11-3

Commercial Sources  
**Combine Answer Sets**  
 Remove Selected  
 Keep Selected

I then selected first 20000 which had the first 20,000 out of 22,214. Also selected “Intersect” option and then “Combine Answer Sets.”

**Combine Answer Sets**

Select saved answer set(s) to combine with your current answer set (20000):

Substance Answer Set Details		Date Saved
<input checked="" type="checkbox"/>	first 20000 (20000) Property "Predicted - Molecular Weight, 441.56-441.64" > substances (22214)	Jul 18, 2012
<input type="checkbox"/>	test (20000) Property "Predicted - Molecular Weight, 441.56-441.64" > substances (22214)	Jul 18, 2012
<input type="checkbox"/>	test2 (20000) Property "Predicted - Molecular Weight, 441.56-441.64" > substances (22214)	Jul 17, 2012
<input type="checkbox"/>	test (20000) Property "Predicted - Molecular Weight, 441.56-441.64" > substances (22214)	Jul 17, 2012
<input type="checkbox"/>	test (20000) Property "Predicted - Molecular Weight, 441.56-441.64" > substances (22214)	Jul 17, 2012

Select an option for combining the answer sets:

**Combine** Include all answers from both sets

**Intersect** Include only answers that appear in both sets

**Exclude** Include only answers from **current answer set (20000)** that are not in **first 20000 (20000)**

**Exclude** Include only answers from **first 20000 (20000)** that are not in **current answer set (20000)**

Combine answer sets and view combined set

**Combine Answer Sets** **Cancel**

I can then now sort the first 20,000 entries since the “Sort by” is not greyed out:

Add KMP Alert Property "Predicted - Molecular Weight, ..." > substances (22214) > Combine S

**Substances** Get References Get Reactions Tools Send to SciPlanner

20000 Substances 0 Selected

Select All Deselect All Sort by: CAS Registry Number

1. Substance Detail  
 1383306-11-3

2. Sub  
 1383

i-Pr

The correct hit was not found in this first 20,000 hits as shown below, so it must be found in the remaining 2,214 hits:

**Substances** Get References Get Reactions Tools Send to SciPlanner

20000 Substances 0 Selected

Select All Deselect All Sort by: Number of References

1. Substance Detail  
 197368-00-6

~27

$$\begin{array}{ccccccc}
 & & \text{O} & & \text{NCO} & & \text{O} \\
 & & || & & | & & || \\
 \text{t-BuO} & - & \text{C} & - \text{CH}_2 - \text{CH}_2 - & \text{C} & - \text{CH}_2 - \text{CH}_2 - & \text{C} - \text{OBu-t} \\
 & & & & | & & \\
 & & & & \text{CH}_2 - \text{CH}_2 - & \text{C} - \text{OBu-t} \\
 & & & & & & || \\
 & & & & & & \text{O}
 \end{array}$$

**C<sub>23</sub> H<sub>39</sub> N O<sub>7</sub>**  
 Heptanedioic acid, 4-[3-(1,1-dimethylethoxy)-3-oxopropyl]-4-isocyanato-, 1,7-bis(1,1-dimethylethyl) ester

So go back to the "bread crumb" menu and reselect the original 22,214 entries:

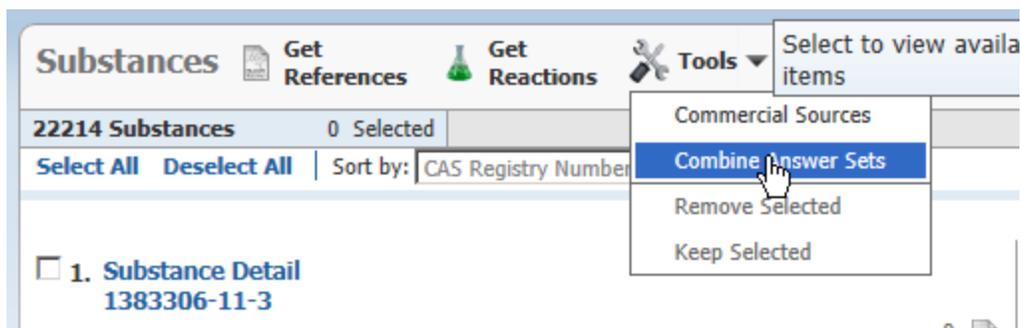
Welcome Jim Little | Sign Out Answer set created with 22214 answers from REGISTRY

Add KMP Alert Property "Predicted - Molecular Weight, ..." > substances (22214) > Combine Substance Answer Sets "first 20000" (20000)

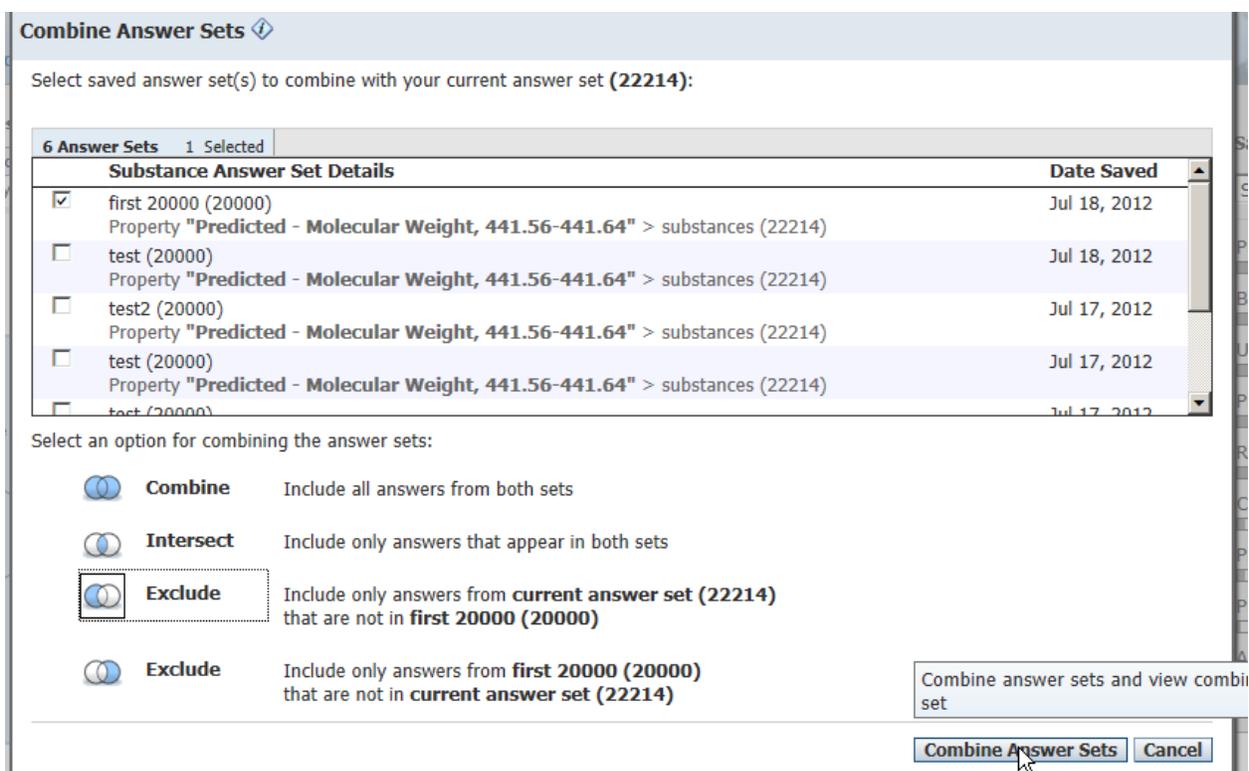
**Substances** Get References Get Reactions Tools Send to SciPlanner

22214 Substances 0 Selected

This will restore the original 22,214 to the current program buffer. Then do Tools/Combine sets again option:



Select first 20000, "Exclude" option, then "Combine Answer Sets."



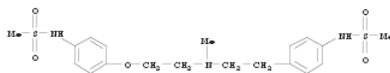
This will return 2,214 that can now be sorted to give the correct answer. However, it is no longer in the first answer set, it not the second one. One would have to determine the correct one by using the fragmentation patter for MS/MS CID spectrum, the isotope ratio, or even the monoisotopic mass instead of the average molecular weight.

Answers per Page [13]

View: [Icons]

1. Substance Detail  
115256-11-6

~618 [Icons]

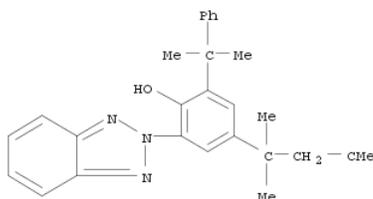


**C<sub>19</sub> H<sub>27</sub> N<sub>3</sub> O<sub>5</sub> S<sub>2</sub>**  
Methanesulfonamide, N-[4-[2-[methyl[2-[4-[(methanesulfonyl)amino]phenoxy]ethyl]amino]ethyl]phenyl]-

Experimental Properties

2. Substance Detail  
73936-91-1

~146 [Icons]



**C<sub>29</sub> H<sub>35</sub> N<sub>3</sub> O**  
Phenol, 2-(2-benzotriazol-2-yl)-6-(1-methyl-1-phenylethyl)-4-(1,1,3,3-tetramethylbutyl)-

3. Substance Detail  
113733-05-0

4. Substance Detail  
110070-51-3

As you can see, this approach is not so easy to use, but it can be accomplished with a little patience. *Normally this would not be a problem with compounds with average molecular weight >600 where it is more difficult to determine one unique molecular formula.*