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.

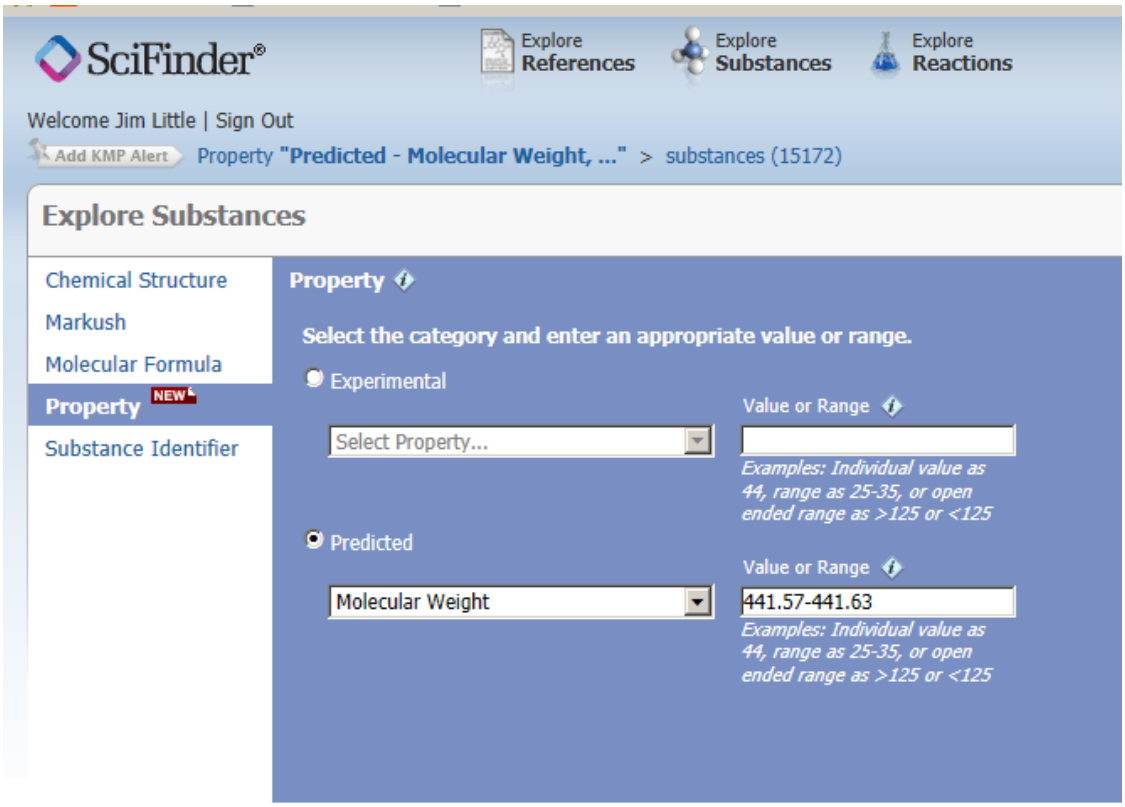
Here is an example that I had used in my initial ASMS article to demonstrate the new capability. Previously, one could only search by average MW in the STN Express interface. Hopefully at some time in the future they will include the capability to search by monoisotopic mass. However, in the meantime,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

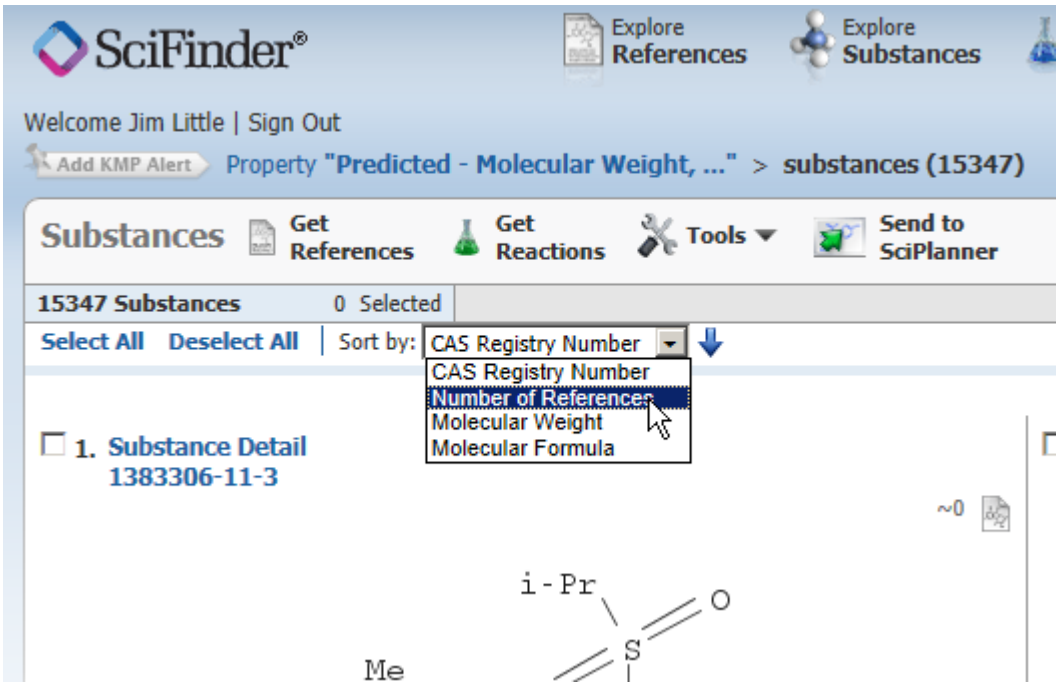
In the article, page 9, I had shown an example for an unknown compound with average MW of 441.60:

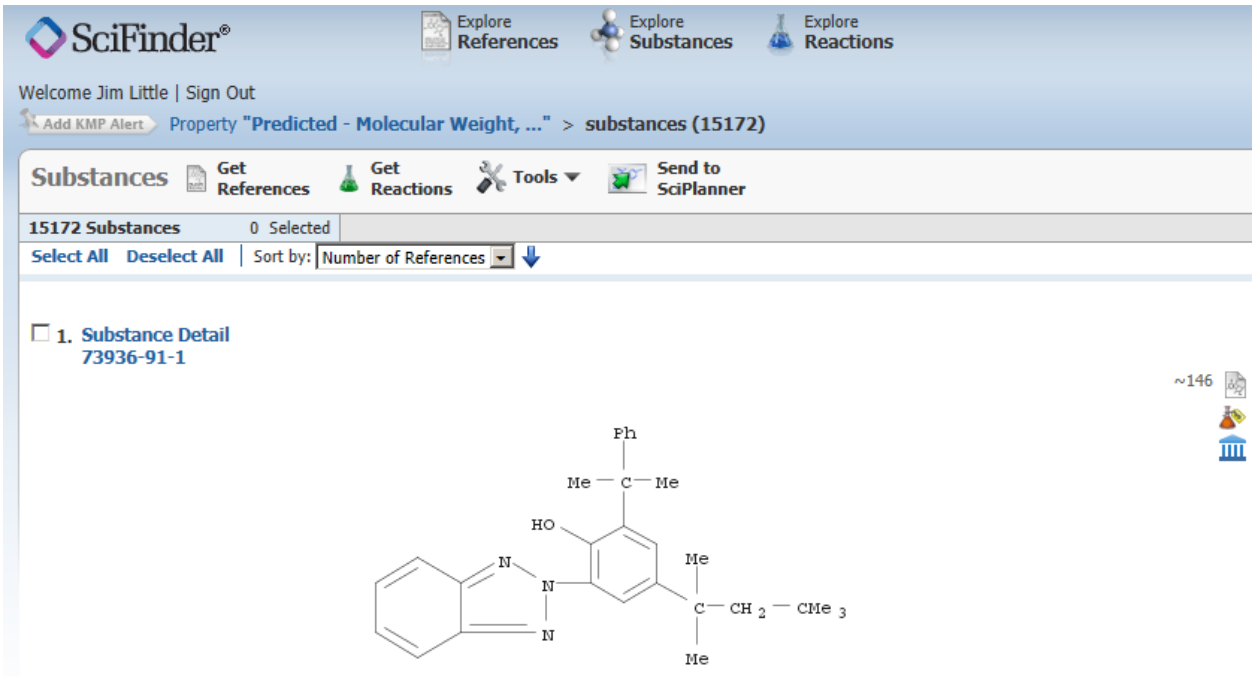
http://www.littledomain.com/james/files/known_unknowns_article.pdf

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



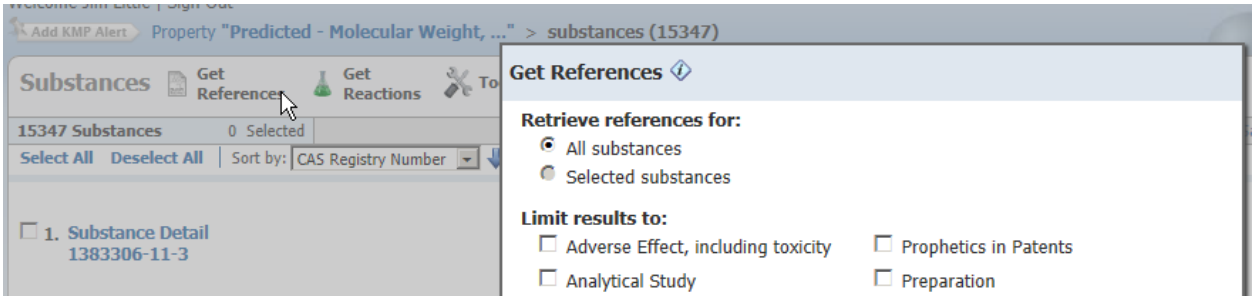
One will get the following display ***after*** sorting by:



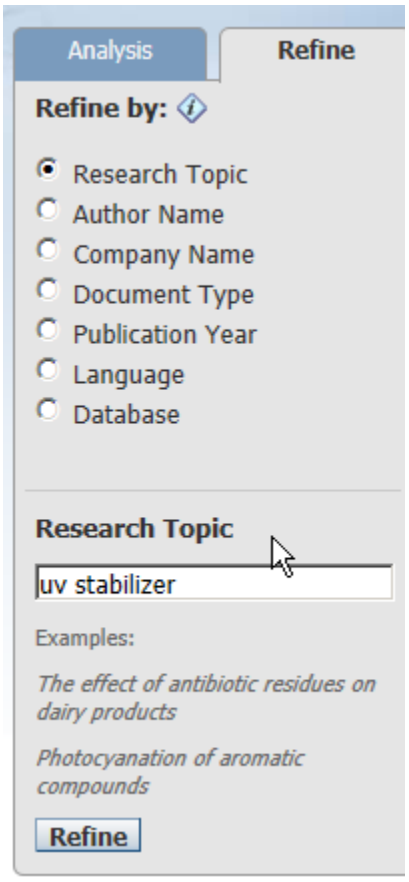


The top hit, CAS No. 73936-91-1 was the correct hit!

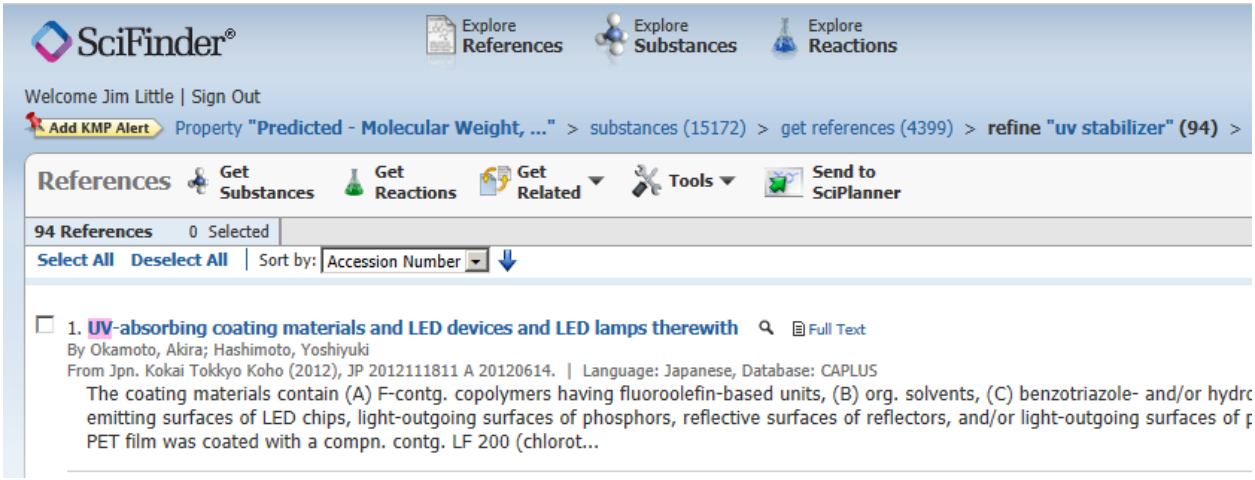
Another alternative for solving this “known unknown” would have been to get all the references:



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:



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



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

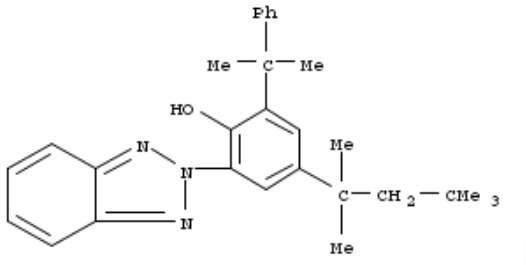
Industrial manufacture; Reactant; Preparation; Reactant or reagent

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

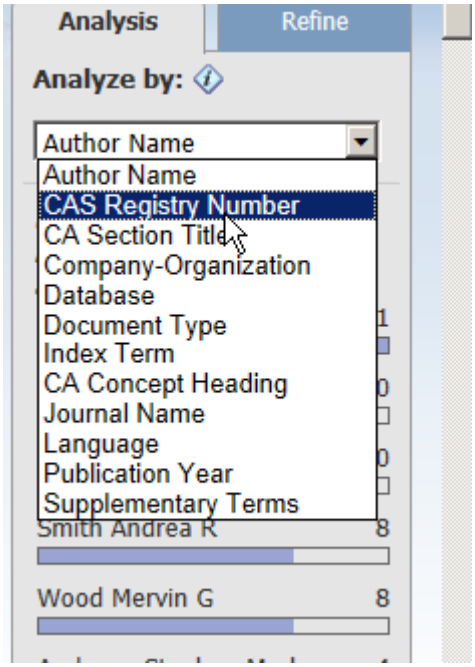
Modifier or additive use; Uses

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

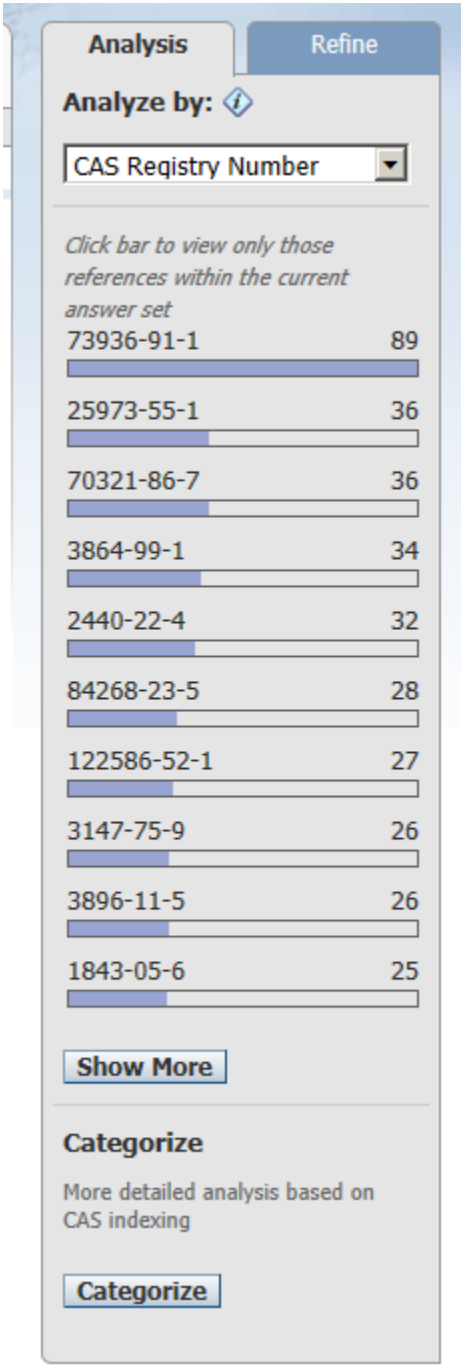


204848-45-3 Tinuvin 479

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:

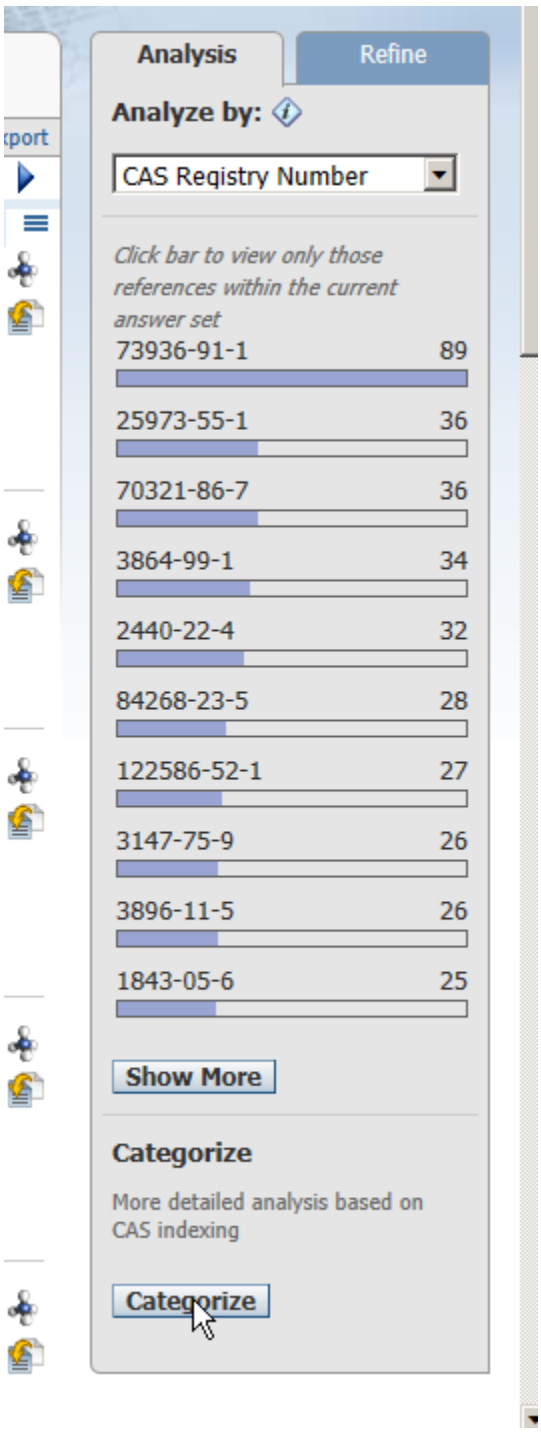


To get the following display:

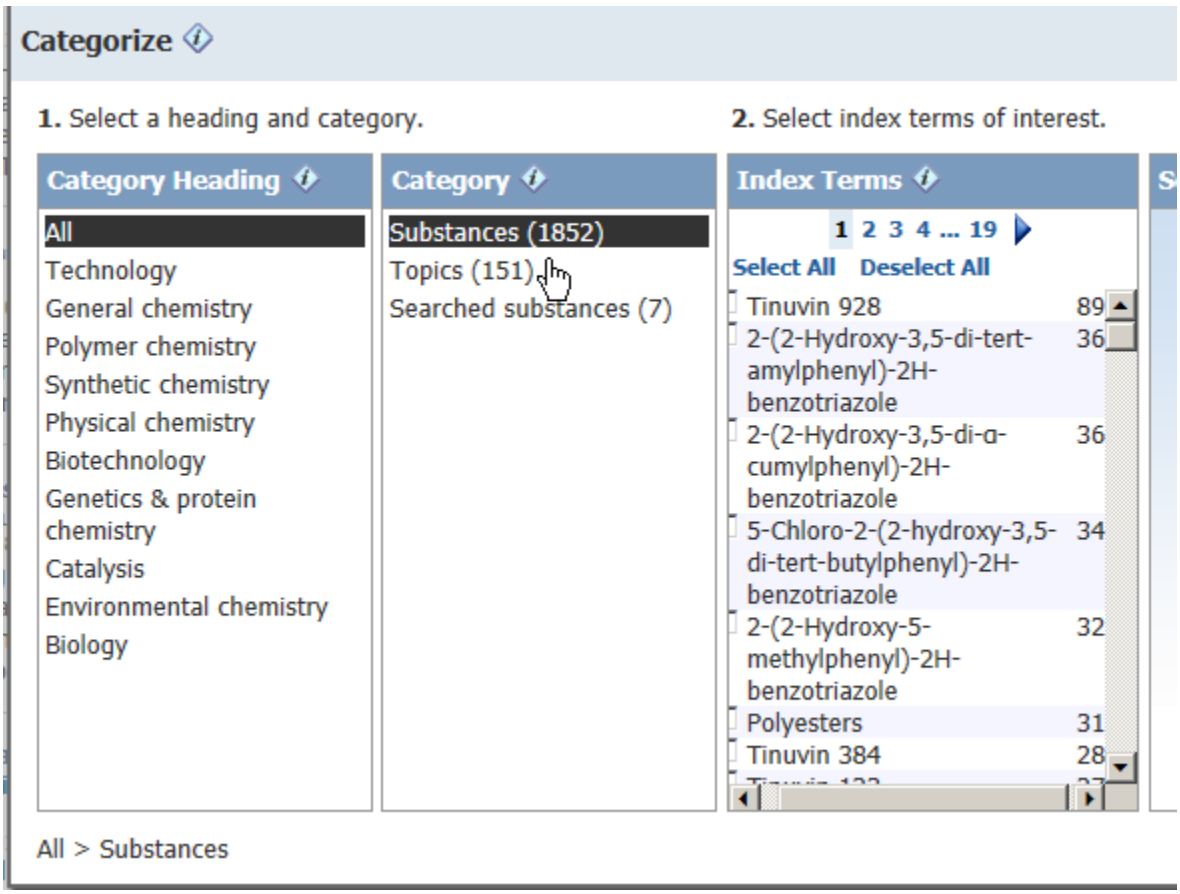


Again the top hit with 89 associated references is the correct compound. This latter display will also list other compounds that tend to be used with this UV stabilizer or in place of this stabilizer.

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



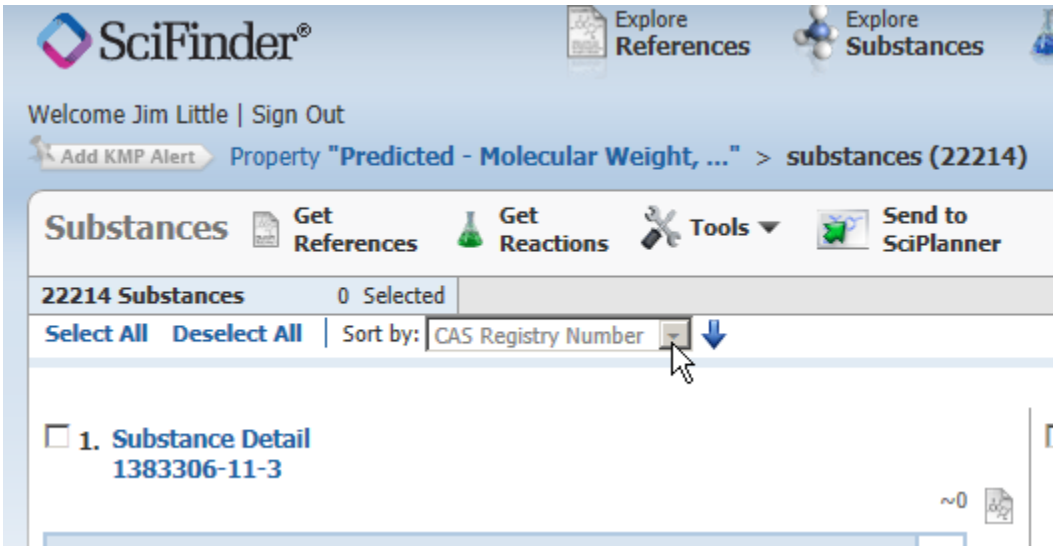
Then select substances:



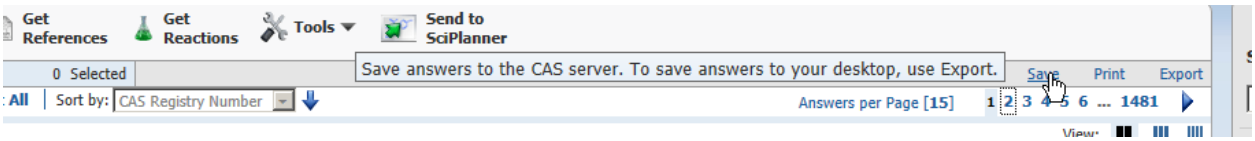
This again has other compounds found in the same references including stabilizers, the type of polymer, etc.

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.

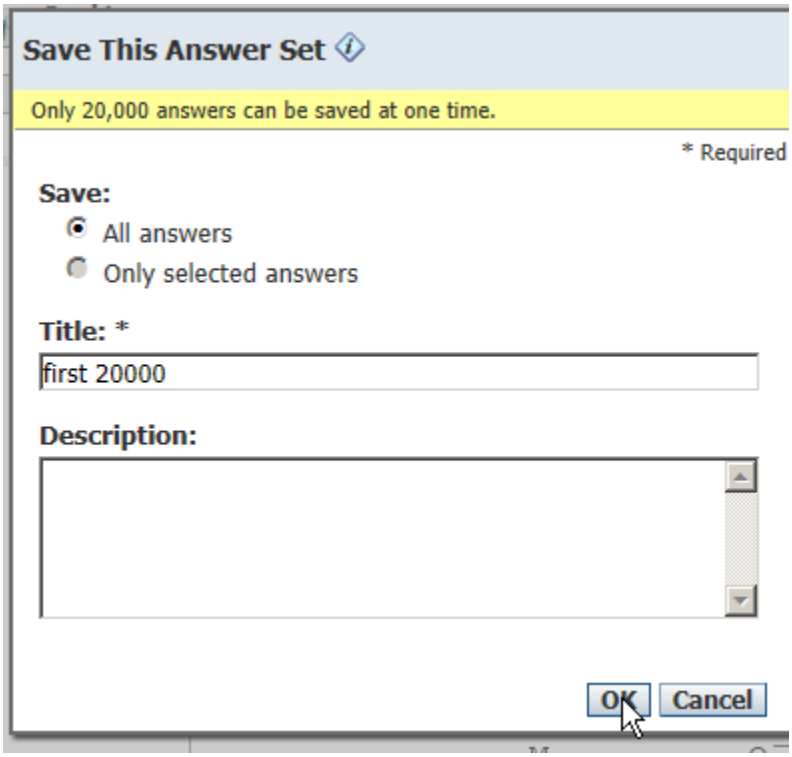
There is a "work-around" to break an answer set that is greater than 20,000 into smaller sizes that can then be sorted by the number of references. *This is not trivial, so be patient.* One must save the answer set. I will demonstrate this with the above example in which I opened the mass window to get >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:



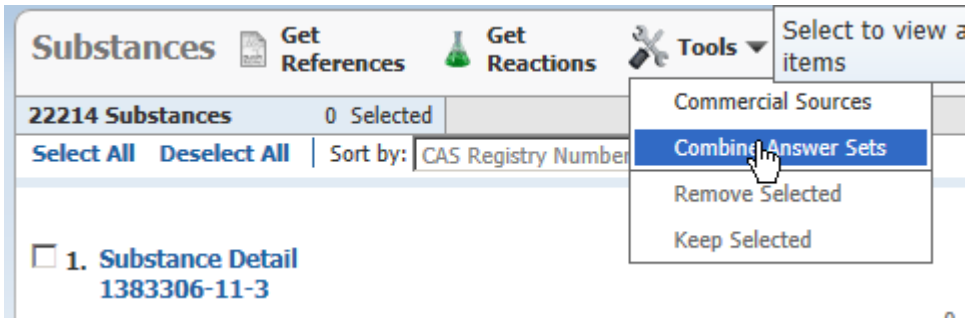
Save this work list (right side of window below):



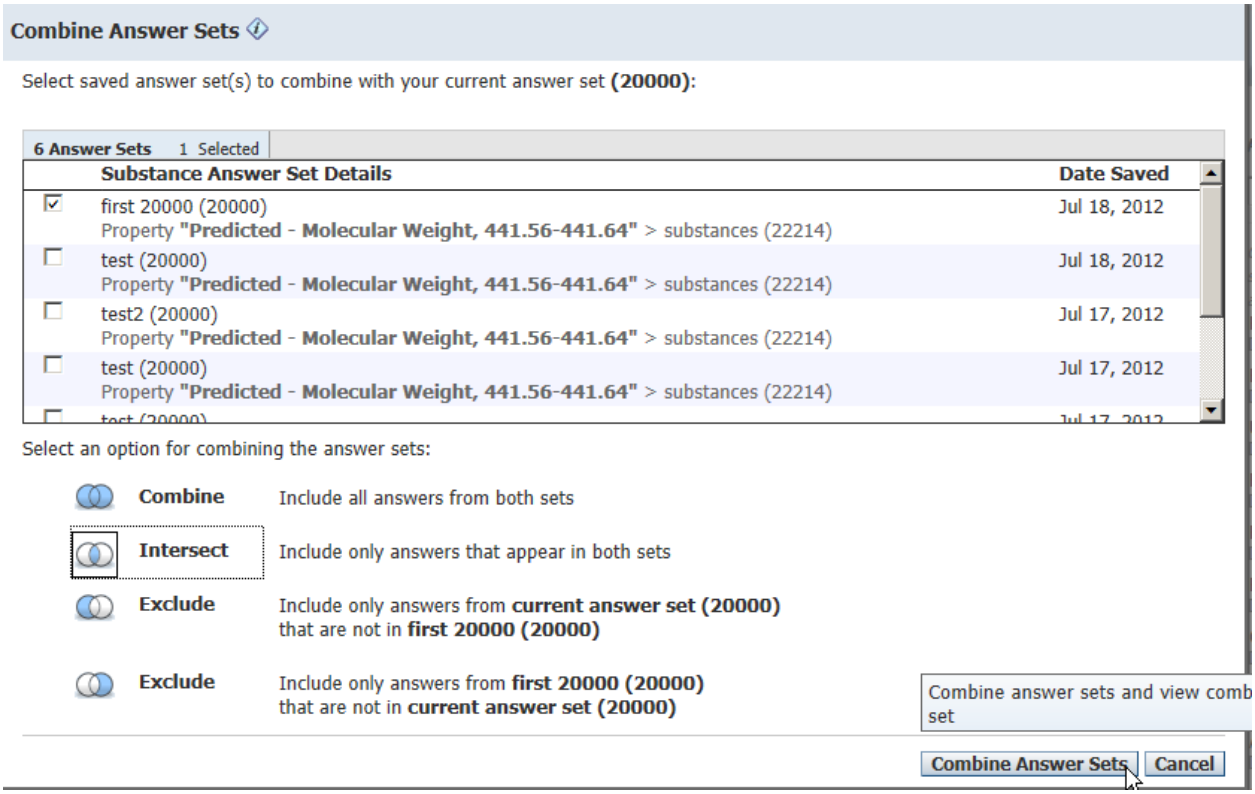
I saved the list as test which will save only the first 20,000 entries of the 22,214 initially obtained:



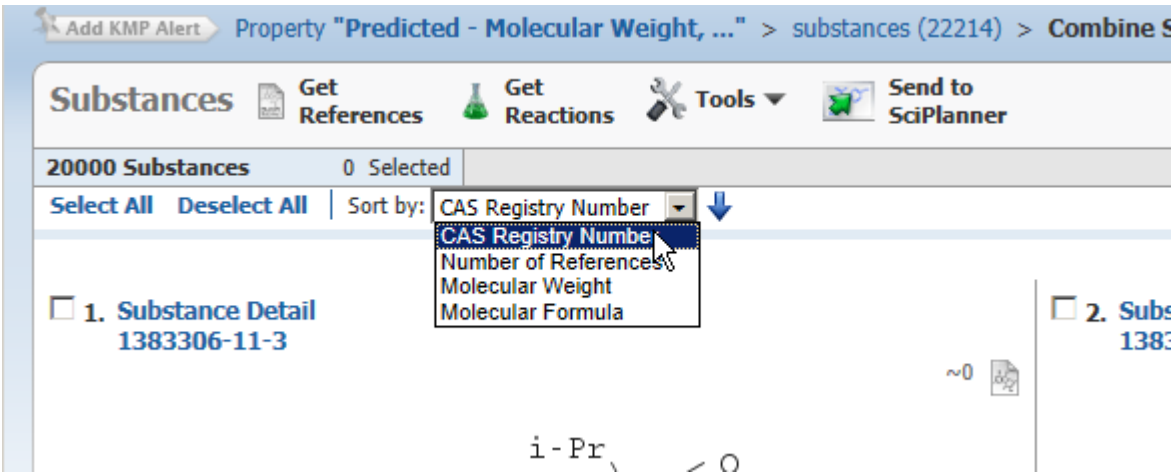
Then use tools/combine answer sets:



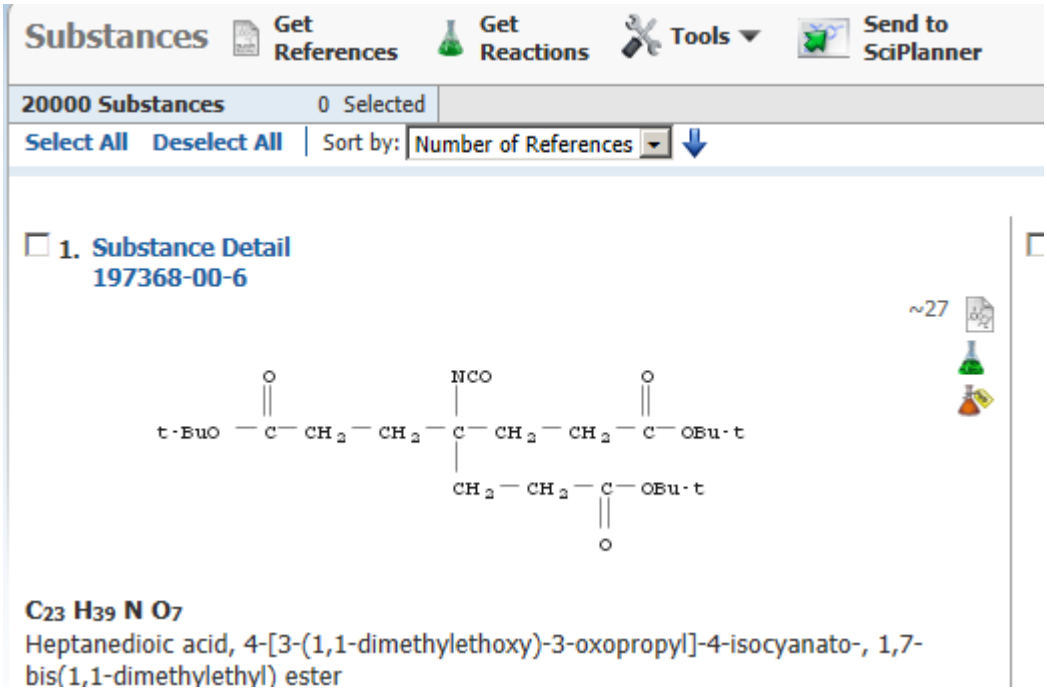
I then selected first 20000 which had the first 20,000 out of 22,214. Also selected "Intersect" option and then "Combine Answer Sets."



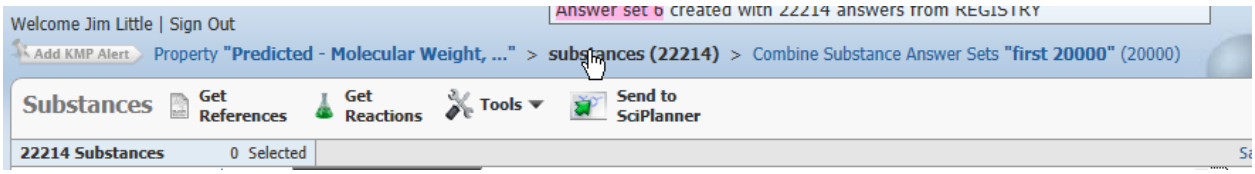
I can then now sort the first 20,000 entries since the "Sort by" is not greyed out:



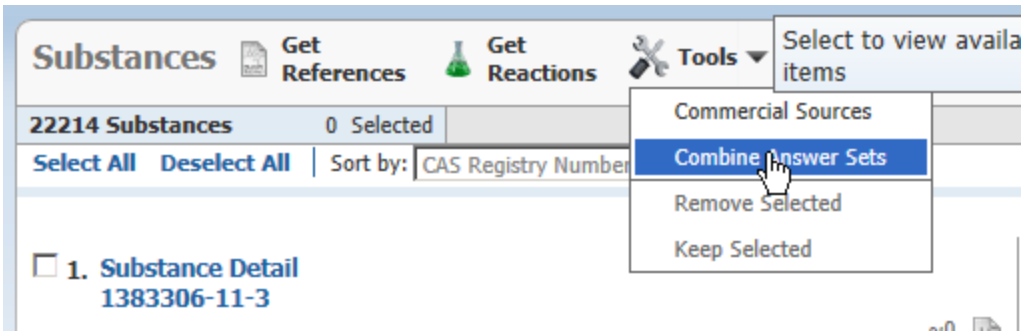
The correct hit was not found in this first 20,000 hits as shown below, so it must be found in the remaining 2,214 hits:



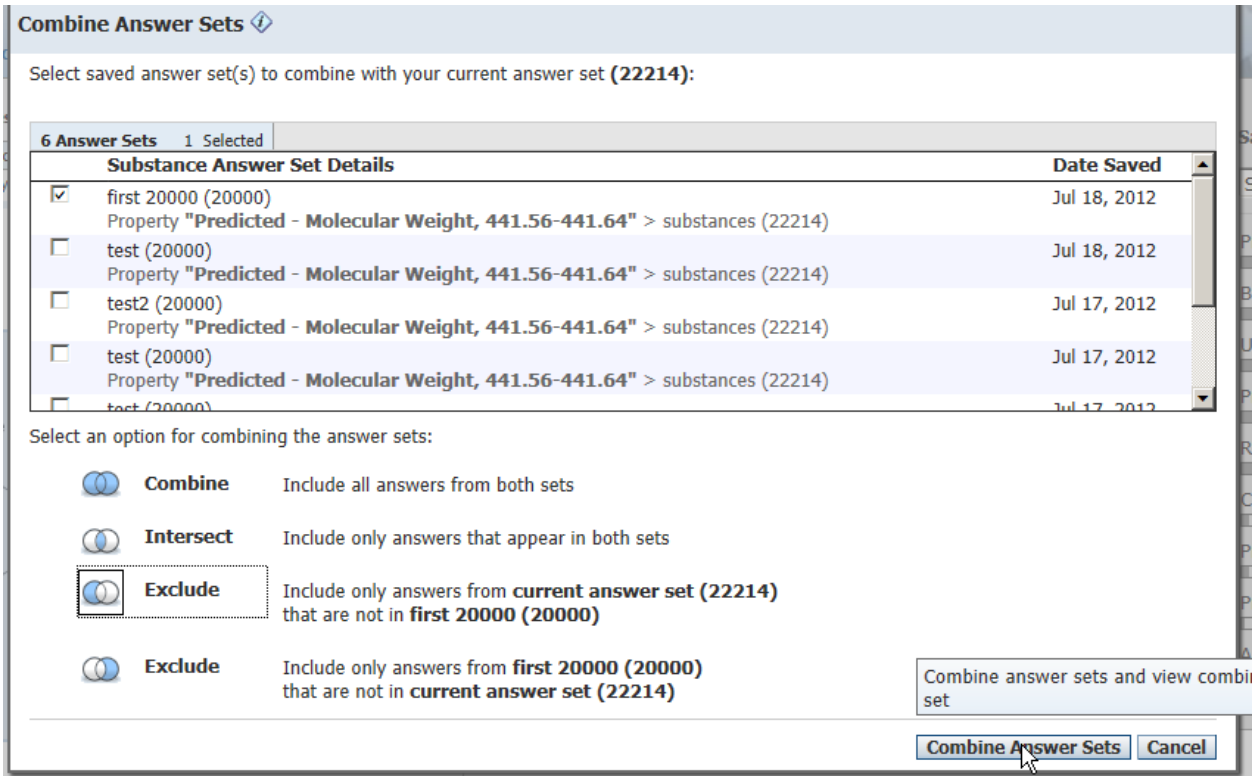
So go back to the “bread crumb” menu and reselect the original 22,214 entries:



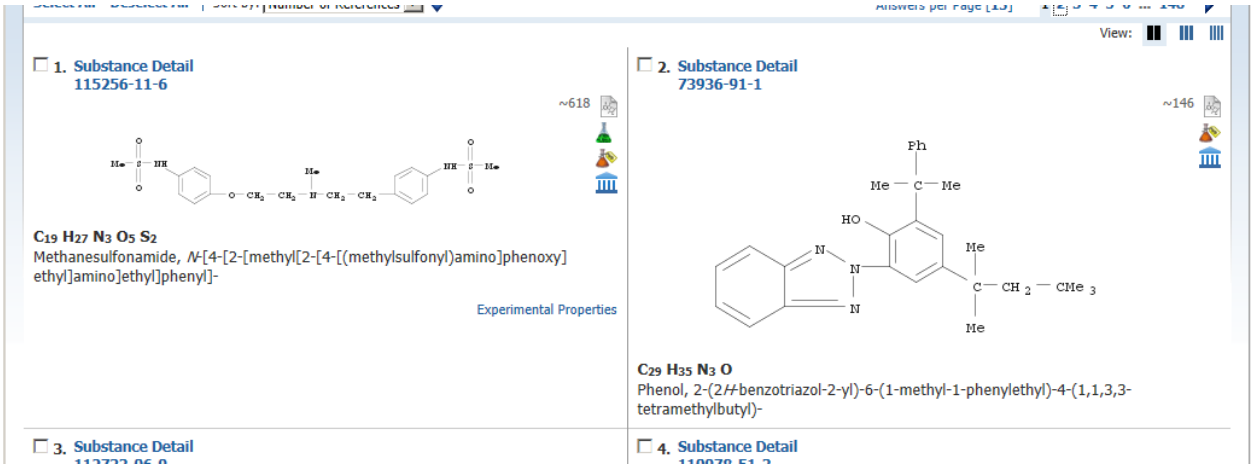
This will restore the original 22,214 to the current program buffer. Then do Tools/Combine sets again option:



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.



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.*