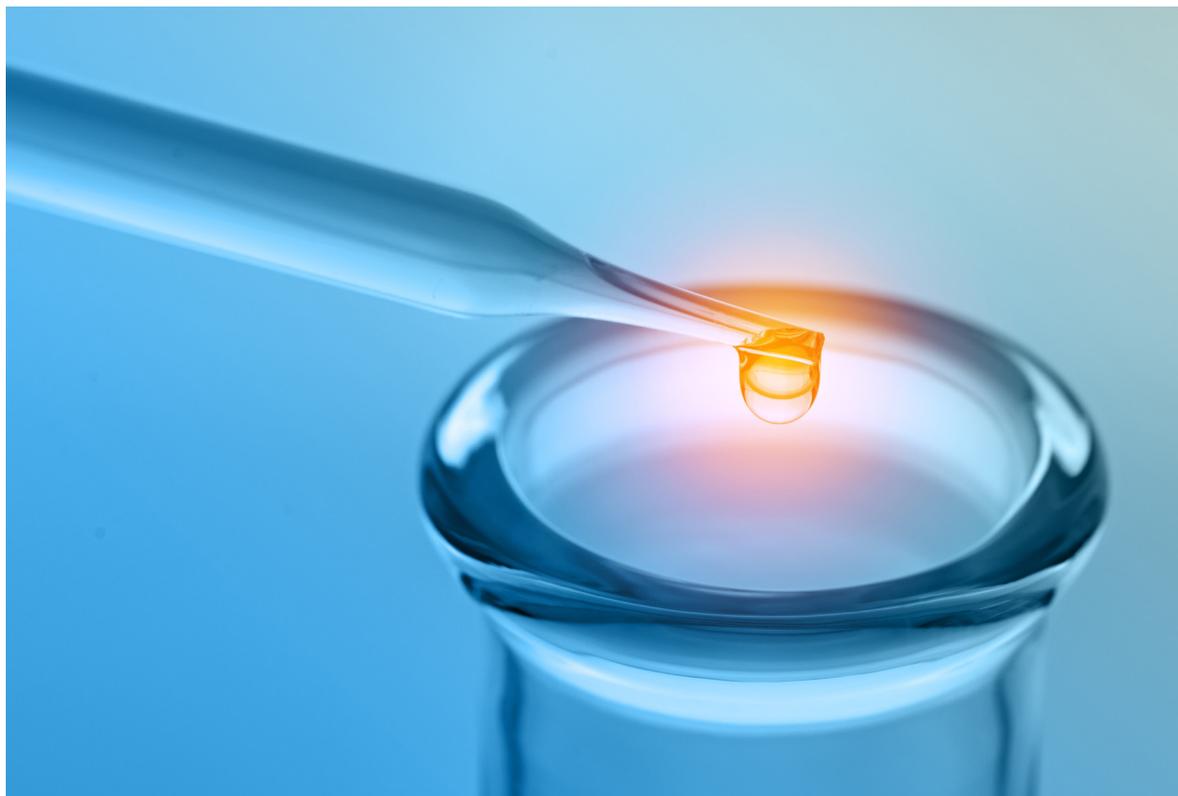


October 2017

OPTIMIZE

SPEXCertiPrep.com/ECS



Non-Target Compound Data Interpretation

by Mark A. Ferry
GC/MS Technical Consultant
MFerry@SPEX.com

A publication by
Environmental Consulting & Supplies



a Division of
SPEXCertiPrep.[®]
Your Science is Our Passion.[®]

Optimize
October 2017

A publication by ECS, a division of SPEX CertiPrep®
203 Norcross Ave.
Metuchen, NJ 08840
Phone: 732-549-7144
Email: crmsales@spex.com
Web: www.spexcertiprep.com/ECS
Copyright © 2017 SPEX CertiPrep®

OPTIMIZE

Non-Target Compound Data Interpretation



Last issue we discussed Target Compound Data Interpretation. In this issue, we will discuss Non-Target Compound Data Interpretation (NIST Library Searches).

First, let's discuss the finer points of the **LSC** and **Lib** menu bars in Enviroquant. I am sure everyone is familiar with the basic procedure to perform library searches but there are a few key points I wish to make. The first is that to improve the quality of your library searches you should do the following under Click Lib\Edit Strategy:

The parameters in the Search Strategy dialog box have been carefully chosen by the Agilent software engineers to give the best search results in most cases. The recommended default values and possible changes are:

U + A = 2. The range is from 1-9. The default is 2. If you are having difficulty obtaining matches increase this setting. It will slow the search but sometimes returns better matches.

Tilting = On. This feature allows the searching algorithm to scale the reference spectrum to fit the unknown spectrum. I suggest leaving it ON as it is in the default mode.

Flag Threshold = 3. The range is 0-99. This is the Threshold setting, so to speak, of the library search. For example, if it is set to 3, then ions below 3% of the base peak won't be taken into consideration. I suggest leaving it at the default of 3%.

Cross-Correlation Sort = Off (box unchecked). By setting this to off, the search algorithm is what's known as PBM match quality (probability based matching) which is the industry standard for EPA methods. I suggest leaving this parameter with default off setting.

Minimum Estimated Purity = 50%. The range is 20-80. The default is 50%. The way this works is as follows: the search algorithm first identifies the base peak in the library match. If that base peak ion is also present at or above the specified % relative to the base peak in the unknown, then the routine reports that match as a possible hit. NOTE: if the base peak in the library match is totally absent from the unknown hit, this feature is deactivated. This is good because if this weren't the case, the disparity caused by differences in scan ranges between what environmental labs and what the contributors of the NIST/NBS/Wiley Library used would make certain matches impossible.

Example:

Unknown spectrum has ion 59 at 100%, ion 45 at 80% and ion 74 at 60%. Possible library search hit has ion 74 at 100%, ion 45 at 30%, ion 59 at 5%, and ion 41 at 95%.

The base peak in the possible library search hit is ion 74. It now checks to see if ion 74 is present in the unknown. It is. It now checks the Minimum Estimated Purity setting. In this case it is set to 50%. Since ion 74 is present at 60% in the unknown, then it passes this test and shows up on the list of possible hits. If the Minimum Estimated Purity was set to 70%, then ion 74's 60% abundance would be BELOW the specified setting and this hit would be ruled out and hence not show up on the list of possible hits.

This may all seem complicated, and indeed it is, so I think that it is best to use the default setting of 50% which is right in the middle of the possible range anyway.

None of the following parameters have any effect on the search itself or on the speed of the search. They all prune entries from the results list after the search is completed.

Low Report MW (molecular weight) = 0.

High Report MW (molecular weight) = 9999.

By setting the MW window from 0-9999 you will not exclude any compounds by virtue of their molecular weight. Realistically, you could use 0-800 and obtain essentially the same results.

Max Hits = 10. This is the number of hits that will be reported. 10 is a good number.

Remove Duplicate CAS Numbers = Off. The only one I disagree with is "Remove Duplicate CAS Numbers" which I always set to ON. The reason for this is that for many compounds, there are numerous entries of its spectra in the NIST/NBS/Wiley Mass Spectral Library. By setting "Remove Duplicate CAS Numbers"=Off, then if an unknown peak is identified in your search, the top 3 or 4 matches will all be the same compound. Well, I don't see any value in that, so by setting to "Remove Duplicate CAS Numbers"=On (i.e. checking the box), then the search only returns the best hit of that compound and then uses the other spots to search for other potential matches. My experience reducing data has been that "Remove Duplicate CAS Numbers"=On works better.

Before I begin my discussion, let me repeat that there are really two “mistakes” that data interpreters/processors should guard against:

- A. False positives - reporting compounds that **aren't** really present
- B. False negatives - failure to report compounds that **are** really present

Should either of these two mistakes occur, all the QC in the world won't atone for it. Sometimes, after what I call “due diligence” (i.e. a thorough, systematic and technically sound approach to data interpretation), two equally competent mass spectroscopists may arrive at different conclusions.

Let me give you a checklist of what to do in qualitative analysis of non-target compounds. But first, before you process any samples you must complete the target compound analysis.

This is because you can't (or shouldn't) do a library search until the system has properly identified the target analytes. If you run a library search prior to having successfully completed target analysis reductions, you might find a target compound show up in your library search. This would be embarrassing to the lab if it showed up in a data package so just remember that library search routines are best performed **only after thorough scrutinization for target compounds has been completed.**

What factors do you take into consideration when determining if it is a hit? I use any and all of the following pieces of information to determine if the compound is truly present. We will assume that when you perform a library search your software identifies the 5 “hits” and ranks them 1-5 based on “Q” value. So, let's begin our discussion with “Q” value.

Q Value

The “Q” value as used in non-target compound analysis is similar, but not quite identical, to its meaning in target compound analysis. The “Q” value as it relates to non-target compound analysis is a reflection of how closely the ion ratios matched in your unknown sample as compared to the compound that the search found.

Keep in mind that “Q” values can be skewed for the following 3 reason:

- A. **Scan range disparity.** We use scan ranges between 35-300 or 550 AMUs; most compounds in the NIST/Wiley Library are typically scanned using 10-800 AMUs. So, even in a “perfect” match, the Q value would be skewed if the compound contains ions outside the 35-300 or 550 scan range. Many volatile compounds have ions in the 10-29 AMU range where this disparity poses a significant problem.

- B. **Tuning disparity.** We tune the mid and high range for passing BFB/DFTPP purposes - the contributors of the NBS/NIST/Wiley Libraries most likely tuned to a different criteria.
- C. **Sample introduction disparity.** The contributors of the NBS/NIST/Wiley Libraries most likely used DIP (direct insertion probe) whereby a few ng of the neat material are injected directly into the mass spec. We introduced our samples into the mass spec via GC (plus a purge and trap in VOA analyses), so the potential for obtaining hits that are breakdown products (as opposed to natural or industrial compounds) is much greater in GC/MS applications.

These 3 factors limit our ability to qualitatively perform non-target analyses. Hence, just because the software ranks the 10 best matches 1-10, you should consider them all equally and evaluate all 10 based on the criteria I will elucidate on the following pages. Think of it as sort of a mystery - the software has identified 10 “suspects” and YOU have to determine which one (if any) is the most likely hit. In spite of the fact that non-target analysis is far less precise than target analysis, there are certain guidelines we can use in our non-target analysis in order to make such a determination. Here are the ones I use:

Absolute Retention Time

Unlike target compound analysis where we have actual known compounds to which we compare retention times, in non-target analysis we have to use our knowledge of gas chromatography and chemistry in order to draw any inferences.

You should ask yourself if this compound *could actually elute* at this RT. Remember non-polar semivolatile DB-5 type columns separate mainly based on boiling point while polar volatile columns like Rtx-502.2 separate mainly based on a combination of boiling point and the analyte affinity for the column. So we can compare the retention time of the unknown hit to the retention time of a known analyte (e.g. the nearest eluting target compound) and determine if that compound could indeed elute at the retention time.

We can also use logical assumptions such as “C16 elutes after C14 after C12” and branched aliphatics/PAHs elute after parent species, etc.

The heavier the molecule and the higher the boiling point, the later in the chromatogram it will generally elute.

So, even though you don't know the absolute retention time of any non-target compounds, you can figure out, to some extent, where in the run it would elute and if the unknown elutes at a significantly different retention time than what you'd expect that would support the absence of that hit. You

should evaluate all 10 hits that the library search report finds for this retention time criteria to see which ones are better candidates than others. **Hits that couldn't possibly elute at the retention time of the unknown should be eliminated from consideration.**

Comparison to Reference Spectrum

This is the most obvious, and oftentimes first, piece of information that the mass spectroscopist evaluates. Which of the 10 hits "looks" most like the unknown hit? It is in this section of the process that we must be mindful of the scan range and tuning disparities mentioned earlier. When "eyeballing" unknown spectra to the ones identified in the search, ask yourself how much better the match could be if the scan ranges were identical. Also, ask yourself how much better the match could be if the systems were tuned identically.

Another important thing to remember is that some of the NBS/NIST/Wiley spectra are what are known as "condensed" spectra because they ONLY contain 10-15 of the major ions. You obviously need to take this into consideration when comparing two spectra. If an ion exists in the unknown spectrum but does not exist in the library's spectrum - that does not necessarily mean that it is not a hit. It could just be that particular ion was excluded in the condensation process. However, if an ion IS present in the library's spectrum but it is absent in the unknown spectrum, that's a much stronger indicator that it's not the correct hit.

Possibility of Coelution

Consider that the peak and resulting spectrum could be a *coelution* of several species. Often the resulting spectrum is an obvious combination of two hits. Other times this isn't so clear. Feel free to call a hit a coelution if that is what your judgment says. Sometimes a TIC (Tentatively Identified Compound) coelutes with a target analyte (you can tell when this has occurred because a large peak on the total ion chromatogram gives a small target compound concentration in your quant report).

Overall Nature of the Sample

Your final piece of information is to consider the overall nature of the sample. Look at the other compounds (both target and non-target) to get a "feel" of the sample. If you are unsure as to whether a hit is truly a target compound - consider the likelihood of this compound being present in this sample. Generally (but not always), compounds alike in chemistry are found together. For example, when you find Benzene, Toluene and Ethyl benzene you find the xylenes. When you find one PAH, you often find others. When you find one halogenated solvent, you often find others. When you find one branched aromatic (e.g. n-propyl benzene), you often find others. When you find one phthalate ester, you often find others, and so on.

Also, consider the source of the sample - was it near a solvent reclamation site? Or an underground storage tank? What kinds of compounds were identified during the target compound analysis portion of data reduction?

Think of the data processing activity as a puzzle - with all the bits of information and tools you have at your disposal as pieces to the puzzle. Individually, no one thing may confirm or deny a hit - but collectively you get a sense of which compound is the right hit - or at least which category of compound it falls into.

Use the following guidelines where applicable:

- *Quantitation* is based on the assumption that the analyte has a RF=1.00 based on the nearest eluting ISTD. This is a potentially grossly inaccurate assumption which renders this quantitation to be very imprecise. For estimation only. Don't get too hung up on the accuracy of the quantitation because without running standards you can't really obtain a very accurate result.
- Many times (actually most of the time) you cannot determine an exact compound but rather a *category* of compound.

KNOW YOUR ESSENTIAL FUNCTIONAL GROUPS: (Note: The "R" groups do not have to be the same)

- Aliphatics (alkanes, alkenes, alkynes). Straight chained or branched.
- Aromatic (Benzene ring structure)
- Aldehydes (R-COH) where R is an aliphatic or aromatic group
- Alcohols (R-C-OH) where R is an aliphatic or aromatic group
- Ethers (R-O-R) where R is an aliphatic or aromatic group
- Esters (R-COO-R) where R is an aliphatic or aromatic group
- Ketones (R-CO-R) where R is an aliphatic or aromatic group
- Polynuclear Aromatic Hydrocarbons (multiple Benzene rings)

Here are some common terms that you can use unless one of the species clearly matches and fits all the above criteria:

- Branched Aromatic (Benzene ring structure with aliphatics)
- Aliphatics (alkanes, alkenes, alkynes). Straight chained or branched
- Aldehydes (R-COH) where R is an aliphatic or aromatic group
- Ethers (R-O-R) where R is an aliphatic or aromatic group

- Esters (R-COO-R) where R is an aliphatic or aromatic group
- Ketones (R-CO-R) where R is an aliphatic or aromatic group
- Aromatic (Benzene ring structure)
- Polynuclear Aromatic Hydrocarbons (multiple Benzene rings)
- Branched Aromatic (Benzene ring structure with aliphatics)
- Phthalate ester
- Possible coelution of _____ and _____ (list the two functional groups)

Look for ion and ion patterns of common functional groups:

GROUP	ION/PATTERNS
Aliphatic Hydrocarbons (alkanes)	43, 57, 71, 85, etc.
Unsaturated Hydrocarbons (alkenes)	41, 55, 69, 83, etc.
Ketones	43, 57, 71
Hydroxyl Groups	45, 59, 73
Aliphatic Amines	44, 58, 72
Phthalate Esters	149
Aromatics	91, 105, 106
Polynuclear Aromatic Hydrocarbons	228-252

Remember the Nitrogen Rule: For compounds containing C, H, O, N and halogens. Molecular ions (a molecular ion is the ion which matches the molecular weight of a compound - but it's not necessarily the base peak in the spectrum) always have an even mass number unless they contain an odd number of nitrogens. Here are two examples of this rule:

Example 1: Benzidine. It has a molecular ion of 184. 184 is an even number and Benzidine has 2 nitrogens, also an even number.

Example 2: Aniline. Aniline has a molecular ion of 93. 93 is an odd number and Aniline has 1 nitrogen, also an odd number.

Consider Stability/Fragmentation Characteristics

Compounds with aromatics rings and double bonds - these are stable molecules which usually produce mass spectra with a large molecular ion and little fragmentation. Molecular ion is the ion with the same mass as the molecule.

Example: Benzene. Molecular weight = 78. Largest ion (called the base peak) is 78. The molecular ion is 78.

Compounds with single bonds, Hydroxyl groups, Ketones and Aldehydes - these are relatively unstable molecules which usually produce mass spectra with a small molecular ion and much fragmentation.

Ions are created based on the characteristics of the molecule plus ionization potentials (we use 70 eV), bond strengths and stability of charged and uncharged species.

Saturated compounds (i.e. compounds without double bonds) give higher than expected secondary, tertiary, etc. ions because of the large fragmenting that occurs.

CAVEAT: Contaminated injection ports and purge and trap systems can create erroneous information so keep your system clean!

FINAL THOUGHT ON THE SUBJECT

As far as I'm concerned, the main thing is to have a systematic, logical, analytically sound approach to non-target compound qualitative analysis. You have a lot less hard data with which to work than in target compound analysis, so interpretive skills really dictate your decision-making process. Even if others may disagree with your opinion, if you've done your "due diligence" and can substantiate your interpretation, then you've done the best you can. If you've considered all of the factors I've given for all 10 of the "suspect" hits and can't conclude anything - just call it "unknown" and let it be.

OPTIMIZE

US Address:

SPEX CertiPrep, Inc.
203 Norcross Avenue
Metuchen, NJ 08840
Tel: 1-800-LAB-SPEX
Fax: 732-603-9647
E-mail: CRMSales@spex.com
Web: www.spexcertiprep.com

UK Address:

SPEX Europe
2 Dalston Gardens
Stanmore
Middlesex, HA7 1BQ UK
Tel: +44 (0) 208 204 6656
Fax: +44 (0) 208 204 6654
E-mail: SPEXEurope@spex.com
Web: www.spexeurope.com



SPEX CertiPrep supports saving our natural resources.



Please Recycle!