LOOKING FOR

New combichem QC standards guard the supply of compounds and ensure HTS results.

BY BING YAN

The dictionary definition of noise is any signal that does not convey useful information and interferes with the desired signal. For example, electrical noise limits the sensitivity of radio receiving systems and, when present at high enough levels, may cause false outputs from digital circuits. The signal-to-noise (S/N) ratio is therefore an important factor when evaluating electronic equipment. In the same way, the S/N ratio of biological screenings is extremely important for the success of drug lead discovery, and it is predominantly determined by the quality of compounds that are screened. Impurities are the main contributors to screening noise.

Besides archive compounds, most chemical supplies currently used for screening come from high-throughput organic synthesis (HTOS) or combinatorial synthesis (*1*). HTOS and combinatorial synthesis are highly effective in producing numerous compounds. However, the quality of compounds made by these methods in the past decade is only average or, in some cases, below average by organic chemistry standards. Three key parameters characterizing the quality of a synthetic product are its identity, purity, and yield (quantity). In traditional organic synthesis, identity is obtained primarily by NMR spectroscopy, purity by elemental analysis, and quantity by the weight of the purified compound. The application of these compounds in biological screenings is therefore unambiguous. However, compound characterization has undergone a dramatic shift in the move from traditional methods to HTOS and combinatorial synthesis.

Because of the limited throughput of traditional analytical methods, a very large number of library compounds, each present in a low quantity, cannot be thoroughly characterized. NMR cannot be used for analyzing every member in large combinatorial libraries because of the slow spectral interpretation process, even though high-throughput spectral acquisition has been achieved (*2*). Elemental analysis is not feasible because of its low throughput and the inadequate sample quantity from combinatorial synthesis. Weighing is cumbersome for compounds synthesized in individual vials and is not feasible for synthesis performed on a multiwell plate. Because of these deficiencies, most combinatorial compounds used in high-throughput screening (HTS) lack some information on identity, purity, and quantity. Using compounds of questionable quality in HTS always generates ambiguous results. Better control and improvement of the purity and yield of desired compounds are of utmost importance in combinatorial chemistry. Because of the various ways to measure the purity, chemistry laboratories have set quite different quality control (QC) criteria before screening.

Rough and ready

In the first QC standard, libraries are not fully characterized before

ILLUSTRATION: SCOTT NEITZKE

QUALITY

HTS. Normally, there are two reasons for not conducting 100% characterization: The throughput of LC/MS is not high enough to analyze the number of compounds found in a large combinatorial library, and the format of one-bead/one-compound in a split-and-pool library cannot be readily analyzed. The practical procedure is to run the screening first and do the purity measurement and structure elucidation only for the positive hits. The structure may eventually be determined by decoding or by NMR and MS analysis when macrobeads are used to make the library.

For libraries containing a large number of compounds, each present in a small quantity, it makes a lot of sense not to carry out a full-blown characterization beforehand. However, for libraries targeted to deliver archived compounds in larger quantities, a complete characterization is necessary.

Slow but sure

In the second QC standard, library compounds are fully characterized by LC/MS, and high-purity compounds are selected for HTS. In this case, LC/MS throughput is the limitation, but with technological innovation, this has become less of a problem. In more and more laboratories, 5000–100,000 compounds are routinely characterized using MUX-LCT eight-channel parallel LC/MS instruments for their identity and purity (*3*).

High-throughput identity determination in combinatorial chemistry has been carried out exclusively by MS. The resolution of a single-quadrupole electrospray ionization (ESI) MS instrument used for analyzing discrete combinatorial libraries is 0.5–1.0 Da. Assuming there is no interference of isobaric impurity ions, such MS measurement may be adequate to address product identification needs. However, this assumption is often invalid in reality. The MUX-LCT system has provided a high-throughput, high-resolution system for compound identification. Researchers at Discovery Partners International (*www.discoverypartners.com*) optimized a parallel high-throughput protocol for accurate mass measurement by using an updated nine-channel multiplexed electrospray LC/UV/time-of-flight (TOF) MS system. With this system, we achieved an accuracy of 10 ppm for 60% of compounds in diverse combinatorial libraries (*4*).

Measurements by UV_{214}, UV_{254}, and evaporative light-scattering detection (ELSD) are also used in-line to determine the purity of each compound. However, these methods measure only the relative amounts of components that respond to a UV_{214}, UV_{254}, or ELSD detector based on the assumption that all substances respond to the detector equally. These detection methods have at least two problems. First, not every component in a sample will respond to a specific detector. Second, the assumption of equal response is invalid.

Our studies showed that the extinction coefficients of organic compounds at 254 nm are more diverse than at 214 nm. Most common impurities are synthetic intermediates or starting materials with a smaller molecular weight (MW) and, likely, less absorbance at 254 nm than the product. Therefore, measurements at UV_{254} may overestimate the purity.

Our results also suggest that the ELSD response factor from diverse compounds is less dependent on their chromophores or structures compared with UV response. However, ELSD also has limitations. On the basis of ELSD and UV_{214} purity determina-

ILLUSTRATION: SCOTT NEITZKE

tions for 100 compounds from 7 different libraries, there is evidently a trend that the relative purity determined by ELSD detection is about 15% higher than that determined at UV_{214}. We observed such an inconsistency in more than 50 combinatorial libraries that we analyzed using both ELSD and UV_{214} detection methods. By studying ELSD responses and quantitation of groups of compounds with different MW and volatility, we found that compounds less than 300 Da generally give a smaller response than expected from their concentration (*5*).

Impurities in the final synthetic products are mostly starting materials, decomposition products, or synthesis intermediates, and they generally have a lower MW than the product. These more-volatile molecules may evaporate with the solvent and result in a smaller response. They may also form liquid droplets instead of solid particles after solvent evaporation. The liquid droplets scatter light poorly and may even absorb light instead of scattering it at certain wavelengths. Even though the ELSD response of these low-MW compounds is small, the compounds usually respond well to UV_{214} detection.

On the basis of these findings, we conclude that UV_{214}, UV_{254}, and ELSD all have limitations. First, the UV and ELSD methods measure only the relative purity. Many impurities, such as trifluoroacetic acid (TFA), inorganic salts, and high-boiling-point solvents, may adversely affect HTS yet are undetectable by these methods. Furthermore, UV_{214} may give incorrect relative purity of the product because of the chromophore variation. Measurement using UV_{214} frequently underestimates the purity of compounds. On the other hand, UV_{254} and ELSD overestimate the relative purity. A more cautious way is to describe the relative purity of a library by both UV_{214} and ELSD.

If the relative purity does not describe the quality of a compound, what will? What is the absolute purity?

The absolute purity (quantitative purity) is the percentage of the desired compound in a sample by weight (*6*). The weight of the desired compound in a sample is obtained by LC/UV, LC/ELSD, LC/chemiluminescence nitrogen detection (CLND), or NMR using a calibration curve generated from a standard.

The relative purity, which is widely used to characterize combinatorial libraries, does not agree with the absolute purity. Figure 1 shows the chromatograms of three compounds and their relative and absolute purity. The relative purity measured by UV_{214} appears to be higher than absolute purity. We also analyzed 50 compounds from 12 different libraries by both the absolute purity method and the LC/UV_{214} method (Figure 2A). The data show that the relative purity is generally 20–40% higher than the absolute purity.

We have seen this inconsistency in almost every compound when both relative and absolute purities were measured. Because the absolute purity is based on weight, these data suggest that about 20–40% of impurities by weight in each sample do not give a sufficient UV_{214} response and are, therefore, undetected or underestimated. Without the absolute-purity determination, the presence of these impurities is practically unknown. These "invisible" impurities can either cause false-positive or false-negative responses in HTS or cause a wrong calculation of experimental concentration and, therefore, wrong assay results. Identifiable invisible impurities include TFA, plastic extracted by organic solvents, inorganic materials, resin washout, reagents, and solvents. Only the absolute purity can adequately describe the real purity of a compound. Thus, this QC standard is outdated.

Inch by inch

In the third QC standard, libraries are fully purified before being characterized by LC/MS, and high-purity compounds are selected for HTS. Invisible impurities can be removed by purification. A purified compound may still show the same relative purity by LC/UV_{214} as an unpurified compound, yet its absolute purity is greatly improved.

We have developed a high-throughput purification (HTP) system to purify all compounds. To test this system, we determined the absolute purity (by LC/CLND and weighing) and relative purity (by LC/UV_{214}) for a randomly selected set of compounds after purification (Figure 2B). The average absolute purity is 91.0%, and the average relative purity is 89.5%. The closeness between the absolute purity and the relative purity has provided compelling evidence that purification is absolutely required for all compounds, even for synthesis products that have a high relative purity (e.g., >90%).

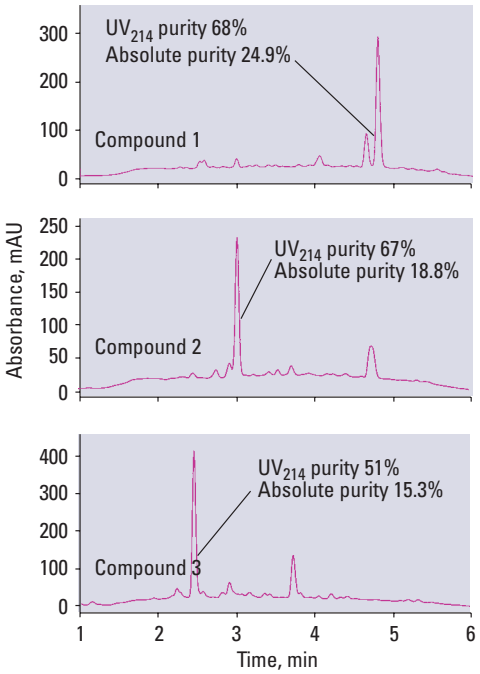

Figure 1. Absolutely relative. Chromatograms for three compounds using UV_{214} detection. The compounds' relative purities are compared with quantitative purity. (Courtesy of Discovery Partners International.)

Establishing a procedure to fully purify all compounds is the first step toward reducing the noise in HTS. However, HTP is not magic, and "garbage in, garbage out" is still true. Our experience is that libraries for purification should be even better optimized than nonpurified libraries for their yield, purity, and "purifiability", the last of which includes factors such as solubility, chromatographic behavior, column-friendly chemicals, and separable impurities. We found that an extra chemical derivatization step is often needed to separate impurities from the product.

When $LC/UV_{214}/MS$ is used as a QC method for combinatorial libraries, we can get only the relative purity of

each compound before going to purification. Although this data contains useful information such as identity, relative purity, and the impurity separability, it is missing the most important information—the quantity. Because of the existence of invisible impurities, the weight of sample and its relative purity bear no relationship to the compound quantity. Three methods for quantifying compound in 96-well plate format are LC/ELSD/MS, LC/CLND/MS, and direct injection quantitative NMR (DI-qNMR) (*2*).

LC/ELSD/MS provides a high-throughput method for obtaining a ballpark estimate of the compound quantity. The average standard deviation is 20–40%, depending on how the calibration curve is generated. Drug molecules and libraries made for drug discovery generally contain nitrogen atoms. Therefore, HPLC with CLND is a high-throughput method for quantitative analysis (*7*). An unrelated compound (usually caffeine) is used to make an external calibration curve. On-line CLND analysis gives quantitative results with a relative error of 10%. Studies have shown that diverse compounds with various numbers of nitrogen atoms yield calibration curves with the same slope on a per-nitrogen basis. This result demonstrated that the method can be used for quantification over a wide range of compounds.

Knowing the quantity of compounds in synthesis wells can give us an idea of how the synthesis reactions went and the amount of desired compounds that we inject onto a preparative HPLC column. If the sample weight is also measured, as in cases where synthesis is done on individual tubes arranged in a 96-well plate format, the absolute purity is obtained.

Another way to determine the absolute purity at a relatively high throughput is qNMR (*8*). This method is based on the fact that the peak areas of a given NMR resonance are directly proportional to the molar amount of that nucleus in the sample. qNMR has several advantages for the analysis of organic compounds: The method is nondestructive; besides the quantitative data, structural (identity) information of the compound is also gathered; and high-throughput spectral-acquisition instruments are commercially available. The main drawback of qNMR is that manual spectral assignment is required. Second, some impurities may not have NMR signals and are therefore undetectable by this technique. qNMR can determine the quantity of a compound or the absolute purity if the weight of the whole sample is determined. Third, we found that the accuracy of the method is reduced if there are impurities in the sample. Our results suggest that peaks from impurities may overlap with the product resonance and erroneously enlarge the product peak area used for quantitation. Therefore, the qNMR method should be used with caution in the analysis of combinatorial library samples.

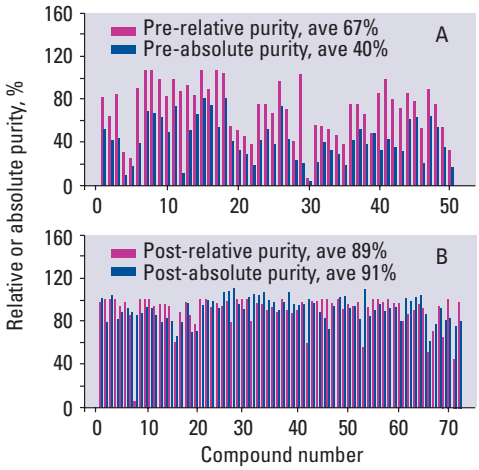

Figure 2. Purely perfect. Comparison of relative purity to quantitative purity for (A) 50 unpurified and (B) 76 purified compounds randomly picked from 14 different libraries. (Courtesy of Discovery Partners International.)

Future directions

The quality of combinatorial libraries determines the success of biological screening in drug discovery programs. Determination of the absolute purity reveals the true library quality and often indicates potential quality problems before full-scale library production. The relative purity can be determined for every member of a large library in a high-throughput mode, but it must be cautiously interpreted. In particular, many impurities are not observable by relative purity measurements using detectors such as UV_{214}, UV_{254}, and ELSD. These invisible impurities may constitute a significant portion of the sample weight. Purification is the only way to remove invisible impurities and improve the absolute purity of a compound, even though it may have a high relative purity before purification.

At one time, "high throughput" meant testing 250–300 samples/instrument/day using fast-gradient HPLC or LC/MS experiments. With the introduction of MUX technology, "high throughput" often means analyzing 2000–2500 samples/instrument/day. The recent introduction of eight-channel UV detectors and four-channel ELSD has provided more parallel detection methods in addition to MUX technology for MS detection. CLND provides a useful quantitative detection method. However, the improvement of its robustness is both important and urgent. Introducing NMR and CLND into high-throughput parallel operation is challenging, yet possible.

On the other hand, several parallel HPLC and CE methods are in development, such as 96-channel CE and 24-channel HPLC, both with UV_{214} detection. But because these methods lack the crucial means to determine the identity of compounds, coupling them with a mass spectrometer has become necessary.

References

(1) Czarnik, A. W.; DeWitt, S. H. *A Practical Guide to Combinatorial Chemistry;* American Chemical Society: Washington, DC, 1997.
(2) Keifer, P. A. *Drugs of the Future* **1998,** *23,* 301.
(3) Fang, L.; et al. *Rapid Commun. Mass Spectr.* **2002,** *16,* 1440–1447.
(4) Fang, L.; et al. *Rapid Commun. Mass Spectr.* **2002,** *17,* 1425–1432.
(5) Fang, L.; Pan, J.; Yan, B. *Bioeng. Biotech. Comb. Chem.* **2001,** *71,* 162–171.
(6) Yan, B.; et al. *J. Comb. Chem.* **2003,** *5,* 547–559.
(7) Fujinari, E. M.; Manes, J. D.; Bizanek, R. *J. Chromatogr. A* **1996,** *743,* 85.
(8) Pinciroli, V.; et al. *J. Comb. Chem.* **2001,** *3,* 434–440.

Bing Yan is director of analytical sciences, Chemistry Division, Discovery Partners International, Inc. (*www.discoverypartners.com*). ■

KEY TERMS: assays and screening, automation, combichem, medicinal chemistry