Chromatography Data Systems in Drug Discovery

The author discusses the roles of mass spectrometry and chromatography data systems within drug discovery analytical laboratories. He also presents ways of using laboratory information management systems to integrate the chemical data.

R.D. McDowall
McDowall Consulting, 73 Murray Avenue, Bromley, Kent BR1 3DJ, United Kingdom

In this article I'll examine the use of chromatography data systems in drug discovery. Of particular interest is analyzing the results of the chemical soups produced by combinatorial chemistry laboratories.

The problem is that chromatographers working in combinatorial chemistry can be in a no-win situation: the combinatorial chemistry increases throughput, and the result is that the number of samples to be assayed also increases. How can this work flow be managed? What can chromatographers do to reduce the waiting times? One idea, presented by Sage and co-workers (1), is to use parallel analysis with an eight-channel multiplexed electrospray interface to increase sample throughput.

General Directions in Drug Discovery

The general trends in drug discovery can be summed up as more, faster, and higher: more compounds synthesized faster than before and with higher purity. This pace will enable the synthesized compounds to be used in high-throughput screening quicker than before, and the samples should last longer than they did previously.

To achieve this process, the chemical identity and purity must be known. Analytical laboratories and chromatographers provide this information.

Analysis in Drug Discovery

The main analytical techniques in drug discovery and their goals are

- high performance liquid chromatography (HPLC), which is used to determine purity using area normalization (percentage of the peak area of total area counts)
- liquid chromatography–mass spectrometry (LC–MS), which is used to determine compound identity based on the prediction from the anticipated compound to be synthesized
- nuclear magnetic resonance spectroscopy, which serves as a backup for compound identification and is used mainly at the start of the library synthesis.

I'll concentrate on HPLC and MS analysis and on the data systems used to manage the data produced. If analysts in small laboratories synthesize more than 100,000 compounds per year, they generally will need to interpret one chromatogram and one spectrum per compound. Still printing on paper? That's a lot of trees.

I discussed this topic in overview in a previous publication (2), but I want to concentrate on the data analysis aspects in more detail. I'll be looking specifically at the chromatography data systems used within laboratories analyzing combinatorial chemistry products.

Analytical data flows in drug discovery: Figure 1 shows the data flows that come through analytical laboratories. The library is planned, starting chemicals are sourced, and the synthesis is planned at the level of the individual 96-well plate. At this point the informatics software can pass information about the expected compound to be synthesized in each plate well. This information is used as the basis for calculating the expected mass ion for the compound in each individual well; the data will be compared with the resulting mass spectrum.

The samples in the plate wells are analyzed first by MS. This step analyzes the products in the solution and confirms that the identity on the premise of the mass ion

is equivalent to the expected molecular weight. Some adduct ions will be produced under certain ionization conditions will require manual interpretation by analysts, but these situations are relatively few in number compared with the numbers of compounds synthesized. For MS analysis, the turnaround time can be limited by the speed of the autosampler feeding the instrument, which will have a wash cycle that could be in the range of 1 min or so. The MS analysis itself can be over within half of that time.

Data handling with MS software is relatively straightforward. The software can produce a report to identify compounds that have been correctly synthesized and those wells that have failed identification.

The next stage is the HPLC gradient analysis. The goal of this step is a short gradient that can separate the compounds used in the synthesis from the anticipated new chemical entities. This process will require some preliminary method development work to ensure the column and mobile-phase conditions are capable of achieving this goal. The gradients are relatively steep and recycled to the starting point relatively quickly. The complication of this step is that it requires 5–6 min.

To save time, analytical chemists ideally should assay only those samples correctly synthesized; that is, the samples that demonstrated the correct mass ion from the MS analysis. There is little point analyzing samples that you already know do not contain your required compound. However, transferring the plate information from the MS data system to the data system running the HPLC can be problematic because these systems usually are manufactured by different vendors and may not talk to each other easily. Due to this communication problem, some laboratories may need to assay the whole plate for purity regardless of the MS results, because it is the easiest way to operate. This process obviously wastes time and effort, so there is a need for the data systems to communicate with each other.

The results of a gradient run can be analyzed using an area percentage normalization method. The total peak area between defined points is used to calculate the percentage purity. This determination is fine if the lambda maximum of the compound is the same as the wavelength used in the detector, but this outcome usually is not the case. Therefore, the calculation usually will have some inaccuracy, but it can be ignored in light of the speed of analysis and turnaround time.

After completing the MS and HPLC analyses, workers must collate the results. If the data systems can communicate effectively, the transfer and collation of results should be relatively easy. As discussed above, they usually don't, so a third application must be used to collate the results and produce a report of the day's analyses, as well as an overall library when the whole analysis has been completed.

Potential solutions with data handling: Many of the problems with data handling in this area have arisen because of the rapid development of combinatorial chemistry and the associated analysis. The introduction of commercial software products has been slow in this area, so many solutions are developed in-house.

The first problem with data handling is that the MS and the chromatography data systems must communicate and transfer data and information between all systems. Few systems can perform these tasks, and they are major requirements to prevent analytical laboratories from becoming larger bottlenecks than they are already. This communication will need a better third-party solution than a spreadsheet. Perhaps one solution can be the use of a commercial laboratory information management system (LIMS).

Figure 2 illustrates how the data systems would communicate via LIMS. The application controlling the synthesis of the combinatorial chemistry library with information about the anticipated identities of compounds to be synthesized would send data to the LIMS. The plate identities and well positions would be identified and downloaded into the MS system. After analysis and interpretation by the MS data system and analysts, the plate details about the wells with positive hits would be uploaded into the LIMS. Wells that were not worth analyzing would be identified and downloaded to the HPLC data system. The LC analysis would skip those wells and assay only those with the appropriate compound present, thus saving time. This approach would save time over the common analyze-everything approach outlined above and would allow the chromatography resources to be used for more added-value analyses.

Is a commercial LIMS worth installing? The answer is yes. The informatics problem in laboratories is relatively simple compared with a quality control laboratory in pharmaceutical manufacturing. Each combinatorial chemistry compound yields two main pieces of information (ignoring repeat analyses): molecular weight (identity) and purity. The reports from the system must be based around this information with a link to the raw data files produced by both instruments.

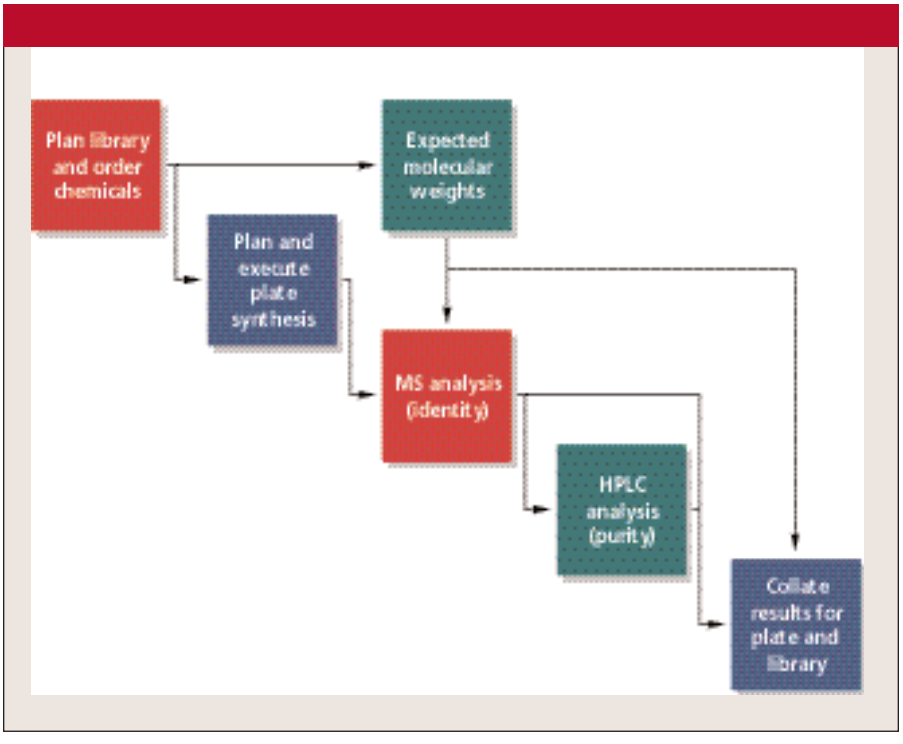

Figure 1: Process flow for analysis of combinatorial chemistry libraries.

Specifications of the compounds are relatively simple: identity and purity compared with the quality control laboratory that may have 10 and 20 analyses with in-house and manufacturing specifications.

The major difference comes from the sheer numbers of compounds that are synthesized and must be managed by the LIMS. The system must be able to cope with the current numbers and be scalable upward by at least an order of magnitude over the next 2–3 years, if expansion continues at the current rate.

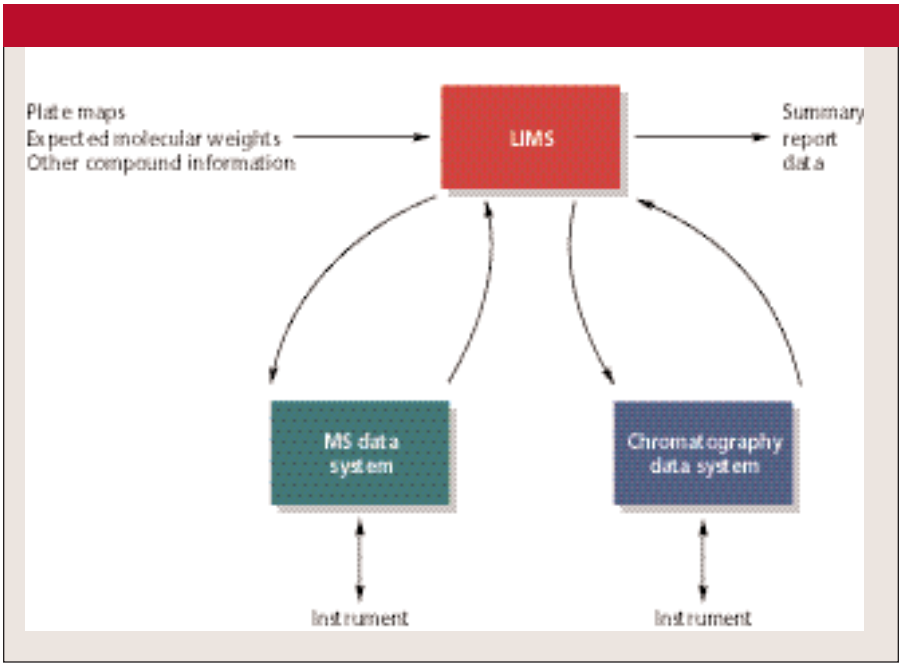

Figure 2: The integration of separate MS and chromatography data systems with a LIMS.

Conclusion

Reducing and, ideally, eliminating the use of paper in drug discovery analytical laboratories can be approached by careful design and integration of the data systems. Usually MS and chromatography data systems are incompatible, but the data from each system can be integrated effectively in a LIMS. The LIMS itself can integrate with other systems that plan the combinatorial chemistry synthesis and plate designs to ensure rapid turnaround of results.

References

(1) A.B. Sage, D. Little, and K. Giles, *LCGC, Current Trends and Developments in Drug Discovery* **18**(5S), S20–S29 (2000).
(2) R.D. McDowall, *LCGC, Current Trends and Developments in Drug Discovery* **18**(5S), S8–S13 (2000). ■