Intelligent Automation of LC–MS Analysis for the Characterization of Compound Libraries

The authors have developed an approach to applying intelligent automation of liquid chromatography–mass spectrometry analysis for analytical screening of compound libraries using Microsoft Visual Basic macros. They performed the initial characterization of compounds using universal analytical methods. Samples that failed to meet selected criteria of analysis were automatically reanalyzed using secondary or alternative analytical conditions. The approach enables automated high-throughput analysis of compounds with diverse chemical properties and does not compromise sample throughput and data quality.
it for the characterization of all samples in a compound library. The same method also can be applied to different libraries. This universal analytical method can be very effective in characterizing most compounds present in a library.

Often, however, the universal analytical method can fail to fully characterize a minority or a significant number of the samples. Several factors account for these failures. Certain compounds might be too polar for the applied separation method, and they will be eluted with the solvent front and prevent the determination of purity. Furthermore, various analytes may be undetected by electrospray MS because of poor ionization efficiency with the selected ionization mode.

The samples that are not fully characterized by the initial attempt using the universal analytical method can be reanalyzed using a secondary or alternative analytical method. However, the time required for an instrument operator to review the results and select appropriate LC–MS methods for each library or sample limits this option when too many compounds fail to meet quality control criteria. A relatively small percentage of failures is accepted as a compromise in exchange for consistently high sample throughput, but these failures can accumulate to a significantly large number when characterizing library collections.

This article reports an approach to applying intelligent automation for high-throughput LC–MS analysis of compound libraries using Microsoft Visual Basic software (Redmond, Washington). Samples that fail to meet selected quality control criteria—for example, an analyte was undetected or eluted with the solvent front—after initial LC–MS analysis with the universal analytical method are automatically selected for repeat analysis with an alternative analytical method. An alternative analytical method is selected for each sample based upon information derived from the target analyte's structure. We will discuss the application of this approach to different instrument platforms.

**Experimental**

We acquired high performance liquid chromatography (HPLC)—grade solvents from Fisher Scientific (Pittsburgh, Pennsylvania). The historic compound libraries were from Nanoscale Combinatorial Synthesis Inc. (Menlo Park, California). The analyte concentrations were approximately 100 μg/mL in 1:1 (v/v) methanol–water. A small amount of dimethyl sulfoxide (approximately 0.2% [v/v]) was present in all samples.

We performed LC separation and UV analysis using an Agilent 1100 HPLC system (Agilent Technologies, Wilmington, Delaware). The injection volume was 2 μL. The LC separation methods used either a 3 cm × 2 mm, 3-μm dp or a 5 cm × 2 mm, 3-μm dp Polaris C-18A analytical column (both from Metachem Technologies, Torrance, California). The mobile phase was water (phase A) and acetonitrile (phase B). Both water and acetonitrile were fortified with 0.1% (v/v) formic acid. Tables I and II list the mobile-phase gradient programs used for each column.

Effluent from the liquid chromatograph was split to allow a flow rate to the ion source of 10–25 μL/min. We performed the analytical characterization of compound libraries using a VG Platform II quadrupole electrospray mass spectrometer (Micromass, Beverly, Massachusetts). Mass analysis was performed using the instrument’s positive-ion mode. The capillary, HV lens, and cone potential were set to 3.3 kV, 0.29 kV, and 27 V, respectively. The source temperature was 100 °C, and the mass range scanned was m/z 200–600. We used Masslynx data-acquisition and Openlynx data-interpretation software applications (both from Micromass) to perform signal acquisition and data interpretation of LC–MS analyses.

Compound library data were incorporates into a worksheet by Accord software (Synopsys Scientific Systems, Leeds, United Kingdom) for Microsoft Excel 97 (Microsoft). Custom automation of data acquisition and data interpretation was accomplished by developing macro programs using Microsoft Visual Basic for Excel 97. Log P data were calculated by Chemdraw Ultra software (Cambridgesoft, Cambridge, Massachusetts) using Crippen's fragmentation (20).

**Results and Discussion**

**Automation of LC–MS analysis:** Workers have various approaches to performing automated LC–MS analysis for analytical screening of compound libraries. The most direct approach is by applying vendor instrument software such as Micromass Masslynx or Thermo Finnigan's X calibur (San Jose, California). Most LC–MS instrument software programs have features that allow automation of data acquisition, interpretation, and reporting. However, many of the features often are inflexible and can present challenges to integrating the LC–MS system with established data infrastructure such as databases, file formats, data networks. Software also relies upon instrument operators for several processes involved in analyses, including run sheet generation, method selection, and custom reporting. In addition, automation features generally are application specific, so they limit custom automated experiments and data interpretation.

Another approach to automated LC–MS analysis that addresses the previously mentioned issues is to implement small software programs or macros that supplement instrument-automation software (16–18, 21). We used a similar approach to bridge data between compound data worksheets and instrument software and acquire and report data. We used Microsoft Visual Basic macros to automate different processes involved in the LC–MS analysis method shown in Figure 1. The Visual Basic macros interacted with the instrument data-acquisition software by send-key command functions; that is, keyboard emulation.

We created compound data worksheets, similar to those shown in Figure 2a,
the development and preparation of compound libraries. The worksheets contained chemistry information — reagents, precursors, and product structures — and data required for the LC–MS analysis — sample location, identification and reference number, and target mass. A Visual Basic program extracted all of the required information from the worksheets and generated a complete run sheet, with each sample assigned a universal analytical method and unique file name (see Figure 3). The run sheet was loaded automatically into the data-acquisition software for analysis. After data acquisition and interpretation of the acquired UV and MS data, the instrument’s data-interpretation software generated a file that summarized the results of the analysis. Selected result information, including target compounds found and relative purity, was extracted directly from the file by a Visual Basic macro and transferred to the starting data worksheet for reporting and archiving purposes. Figure 2b shows an example of the worksheet report format.

Visual Basic macros can complement instrument automation software. Worksheet information can be bridged efficiently to data-acquisition software to facilitate custom automation of LC–MS analysis. This macro-enabled LC–MS analysis relies less upon instrument operators, and it could be used as a foundation for incorporating the intelligent feature into the automated analysis.

**Automated intelligent LC–MS method selection:** When working with large numbers of samples, analysts commonly use a single universal LC–MS analysis method to characterize all samples from various libraries. Because compound libraries often comprise analytes with diverse physical properties, it is common to encounter analytical failures in which several samples analyzed using the universal method were not fully characterized or failed to meet selected quality control criteria; for example, an analyte was not detected or its purity could not be determined. Repeating analyses using secondary methods is a time-consuming process that requires instrument operators to review results, assign appropriate analytical methods, and generate new run sheets. As a result, workers repeat analyses only when they encounter a significant number of failures. A relatively small number of analysis failures often are acceptable to maintain consistent sample throughput.

To minimize the number of failures, Visual Basic macros should be implemented to automatically repeat the analysis of failed samples. Following the initial analysis of compound libraries using the universal analysis method, the macro opens the custom report data file from the analysis (Figure 2b) and searches for samples that fail quality control criteria. The macro uses the general information of the failed samples — sample location, molecular weight, and reference number — to create a new run sheet (Figure 1) that will be used for a subsequent analysis.

The macro also can examine and extract additional information present in the compound data worksheet. The information can be entered into a data worksheet by a user or derived directly from the analyte structure. The macro can use this additional information to intelligently select the most appropriate analytical method for each

---

**Figure 1:** Schematic for the LC–MS analysis process enabled by Visual Basic macros.
failed sample from the initial analysis. One property that can be calculated from the analyte structure is the octanol–water partition coefficient (log $P$). Log $P$ typically correlates to the reversed-phase retention properties of an analyte, and the macros can use it as a reference for separation method selection.

Figures 4a and 4b show the UV chromatograms from an analysis of selected compounds from a historical library using the LC primary separation program shown in Table I. Compounds with a calculated log $P$ value of roughly 2 readily were eluted beyond the solvent front peak (approximately 0.1 min), caused primarily by the presence of dimethyl sulfoxide. Separating the target analyte from the solvent front allowed the determination of relative UV purity information, as shown in Figure 4a.

The primary separation method, which used a 3-cm long LC column and a 3-min total analysis time, is well suited for fast characterization of large sample arrays and may be implemented as part of the univer-

---

**Figure 2:** Screen captures of (a) a compound library development worksheet and (b) report output.

---

**Figure 3:** Screen captures of Visual Basic macros interface.

---

**Table I:**

<table>
<thead>
<tr>
<th>Column 1</th>
<th>Column 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound A</td>
<td>log $P$ 2.1</td>
</tr>
<tr>
<td>Compound B</td>
<td>log $P$ 1.9</td>
</tr>
<tr>
<td>Compound C</td>
<td>log $P$ 2.3</td>
</tr>
<tr>
<td>Compound D</td>
<td>log $P$ 2.0</td>
</tr>
</tbody>
</table>

---

**Figure 4a:** UV chromatogram showing elution of compounds with log $P$ values of 2.1 and 2.0.

---

**Figure 4b:** UV chromatogram showing elution of compounds with log $P$ values of 1.9 and 2.3.
Figure 4: UV chromatograms for the analysis of compounds with (a) calculated log P = 2 using the primary LC separation method, (b) calculated log P = 1 using the primary LC method, and (c) calculated log P = 1 using the secondary LC separation method.

Figure 5 illustrates the process used by the macro to select an analytical separation method for each analyte after an analysis using the universal method. The macro used calculated log P values as a reference for LC method selection. For failed compounds that have relatively low calculated log P values, the Visual Basic program selected an analytical method more appropriate for polar compounds; for example, a method that used a slow gradient and a longer column. Figure 4b shows the UV chromatogram obtained for the same compound (calculated log P of 1) as shown in Figure 4b using the secondary LC conditions (Table II). Using a slower gradient and longer LC column separated the analyte peak from the dimethyl sulfoxide peak and enabled the determination of compound UV purity.

We used log P as a reference for separation method selection when characterizing a historic compound library of 80 samples. The primary universal method (Table I) was sufficient to characterize approximately 94% of the samples (75 samples). However, the relative purity of approximately 6% of the analytes (five samples) could not be determined because of poor LC retention. The calculated log P values of the failed analytes were relatively low (less than approximately 1.5). By repeating the analysis of the failed compounds using the secondary separation method (Table II), approximately 60% of the failed analytes were sufficiently retained on the column to permit relative purity determination. Although the samples and results from this analysis may not be representative of all libraries, it illustrates how structural information can be used for automated LC-MS analysis and method selection.

Using the target structural information for automated LC-MS method selection and analysis can improve the efficiency of compound library characterization and increase the number of successful analyses. Furthermore, the use of automated intelligent method selection for failed analyses could allow faster universal separation.

In addition to column and mobile-phase gradient selection, several method parameters can be controlled or selected by the Visual Basic program, including: gradient selection, several method parameters can be controlled or selected by the Visual Basic program, including: gradient selection, spray potential, cone potential, ionization mode, UV detector wavelength, mobile-phase composition, and mass range scanned. Each of these parameters can be controlled directly by selecting various analytical method files during the run sheet-generation process. As a result, other information can be added to the data worksheet or derived from the analyte structure that potentially can be used by the Visual Basic macros to optimize the various instrument parameters selectively for the intended target analyte. The acid dissociation constant (pKa), which is based upon structural information or simple functional group detection algorithms, can be useful in the selection of mobile-phase composition (acid or base additives) and ionization mode. Data derived from the analyte structure also can be useful in selecting an appropriate UV wavelength for an analyte. Chromatographers also can develop algorithms to match the structure of the target analyte with pre-
Flow chart for the intelligent selection of LC–MS analytical methods.

**Figure 5:** Flow chart for the intelligent selection of LC–MS analytical methods.