Ensuring the Quality of Registered Compounds in a Drug Discovery Environment – A Multidisciplinary Approach

Ryan Sasaki, Tara Sinclair
Advanced Chemistry Development, Inc.,

Overview

Since 2007, Lexicon Pharmaceuticals have demonstrated that a practical automated verification system using HPLC, LC/MS, and 1D and 2D NMR can be implemented in an industrial/pharmaceutical environment. This system has proven to be robust, and provides added value to compound collection integrity and quality. In addition, it has provided a means to apply analytical review, with minimal staffing, to entire collections of compounds which would otherwise not be possible. The system identifies a manageable subset of the entire compound collection for an analytical specialist, eliminating the need to look at compounds with a high degree of confidence and concentrate on those more likely to be of issue. Simplicity has been the key to its effective and growing use in the organization. The benefits of this system to synthetic chemists are clear—the amount of additional effort on behalf of the chemists' workflow is negligible but provides a plethora of useful information, including an overall assessment of their compound tentative assignments. With this combination, user acceptance and trust from the medicinal chemistry community was gained. This poster will focus mainly on the addition of an NMR verification filter using commercial NMR software to existing automated in-house HPLC and LC/MS assessment tools.

Introduction

Based on the evolution and innovation of analytical instrumentation in the "open-access" movement, today, the majority of the routine analysis is done by a chemist and most samples are being prepared and developed without passing the analytical specialist's eyes. While this evolution has certainly improved on sample turnaround times, the new burden is now on the synthetic chemists who have to process and interpret their own analytical data. With analytical specialists no longer reviewing all samples, the risk of having erroneous samples is higher.

Combined with this risk, is the demand to increase the number of compounds produced by the chemists. Couple this with the ability to generate even greater volumes of analytical data, and the analytical interpretation aspects have struggled to keep pace with compound reviewing demands. For these reasons, automated structure verification using software has been investigated in this area to improve throughput and QC of registration libraries.

How it Works

Typically compounds submitted to corporate registration systems are assumed to be relatively pure and conform to the submitted structure with a reliance on the mass spectrum confirming just the formula weight. This presents a problem for regioisomers, potentially derived from alternative positions of substitution in a reaction, which are generally not well characterized by classical MS analysis. In many cases a greater dependence on NMR is usually necessary. The routine experiment of choice for synthetic chemists is ^{1}H NMR. Unfortunately the verification system in place is NMR prediction based, and ^{1}H NMR prediction alone presents several limitations that prevent its usefulness for routine "qualifying" of a compound, especially in the case of regioisomers.

In NMR, the HSQC experiment provides ^{13}C chemical shift information that would otherwise be unavailable without the larger quantities of compound needed to collect a 1D ^{13}C spectrum generated in the later stages of drug discovery. The existing compound registration and submission system interface has been modified to show a simple traffic light result of red, yellow, or green for each "qualifying" analysis that allows for rapid visualization of the overall result.

ACD/Labs
Advanced Chemistry Development, Inc.
Tel: (416) 368-3435
Fax: (416) 368-5596
Toll Free: 1-800-304-3988
Email: info@acdlabs.com
www.acdlabs.com

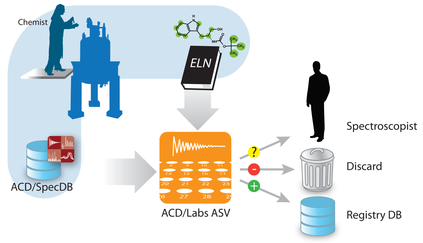
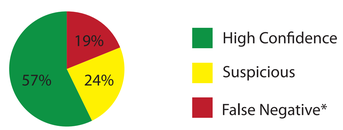


Figure 1. Workflow schematic of a multidisciplinary automated verification/structure confirmation approach.

Results

Since initial implementation, a percentage of the total possible number of compounds submitted have undergone automated verification and to date, over 2000 compound data sets have been evaluated. A key metric for determining the accuracy of the system is the reporting of false negatives. Each negative result is carefully inspected by an analytical specialist to determine whether it is truly a negative result or a false negative, and the cause. True negatives are often due to a structure drawing error by the chemist at registration, and are either corrected or excluded from the inventory when they fail for other reasons. False negatives are recorded and the information is used to improve the system on an ongoing basis. As an aggregate, a false negative rate of 19% across MedChem and intermediate compounds evaluated was reported (Figure 2).



* as determined by subsequent inspection

Figure 2. System verification results for 1402 MedChem and Intermediate compound data sets.

Chemists are presented with an automatically generated report (Figure 3) that shows a tentative assignment of their compound. This in turn streamlines NMR interpretation and generates a reasonable proposed starting point for assignment.

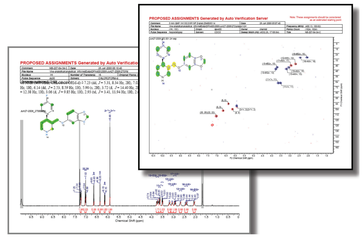


Figure 3. Automatically generated assignment report with color coded indicators for assignment confidence.

Validation

To validate the system, 152 benchmark commercial chemicals were used to challenge and optimize the system using a series of positive and negative controls (correct and incorrect structures). The ultimate goal is to identify evaluation criteria which effectively balance the risk of false positives against the management of a reasonable subset of data for review (false negatives). This benchmark study includes a series of somewhat simpler sets of compounds for positive controls relative to medicinal chemistry submissions. On the other hand, to have confidence in the diagnostic ability of the software to avoid false positives, a challenging set of negative control compounds (relative to the obvious errors we generally see) is required.

For this benchmark, a false negative rate of 8% was observed while fixing evaluation criteria and thresholds to keep the false positive rate near or less than 20% (Figure 4). This is a critical step in establishing the basis for organizational confidence in the results.

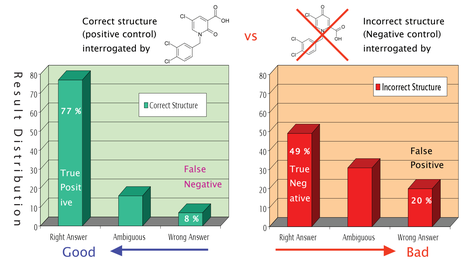


Figure 4. Benchmark of 152 commercial reference compound data sets used to challenge the auto verification system.

Benefits

To date over 2000 sets of compounds evaluated under the automated system using ^{1}H, HSQC, LC/MS, and HPLC data have helped improve the overall quality of our registration database. These data sets have also allowed us to examine the strengths and weaknesses of the current implementation so that regular improvements can be made.

The benefits of this system are numerous and include a reduction in input needed from skilled NMR Spectroscopists, allowing them to focus their limited time on compounds potentially needing more scrutiny.

In addition to time savings and staffing considerations, other benefits include greater accuracy of resulting structural information leading to improved interpretability of Structure Activity Relationship (SAR) information for lead optimization of *in vitro* and *in vivo* screening results. Furthermore, the long term benefit of adding only correct structures/compounds to our screening library prevents future errors that can lead to missing or mistaking tractable hits for a future lead optimization campaign.

References

1. P. Keyes, G. Hernadez, G. Cianchetta, J. Robinson, B Lefebvre, *Magn. Reson. Chem.* **2009**, 47, 38-52.
2. Oct 28, 2008, European Users' Meeting *Validating Compound Registrations with Automated NMR Verification in Open-Access.*