LC TROUBLESHOOTING

Calibration Problems — A Case Study

Unexpected results from calibration standards create confusion in a clinical liquid chromatography (LC) method.

Recently, I received an inquiry from a reader regarding a problem he encountered with a routine liquid chromatography (LC) method in his clinical laboratory. He had prepared a fresh calibration standard (check sample) for the analyte of interest (I'll call it "X" to keep the reader's laboratory anonymous), yet when he assayed a blank sample spiked with 160 ppm of X, he found an indicated 400 ppm. This was puzzling and not a problem normally encountered, so he sent the sample to another laboratory that was analyzing the same compound by gas chromatography (GC), and their results showed that the spiked sample indeed contained 160 ppm of X. At this point he contacted me to help figure out what was happening. As we look at possible causes for and solutions to this problem, we can use this as a specific example to which we can apply general troubleshooting principles.

Background

Before we get further, let's take a look at the method, which is designed for the analysis of X in serum. Samples are prepared by taking an aliquot of serum, adding an aliquot of internal standard (IS), and a small amount of hydrochloric acid to acidify it. The solution is vortexed to mix, then an aliquot of dichloromethane is added, the solution is vortexed again, and then centrifuged to separate the two phases. The dichloromethane phase is removed, evaporated to dryness, and reconstituted in the injection solvent. The separation conditions comprise a reversed-phase column (size, stationary phase, and flow rate were not mentioned) with an isocratic mobile phase

of acetonitrile, water, and trifluoroacetic acid. Ultraviolet (UV) detection is used. The chromatographic conditions give typical retention times of 9 min (IS) and 12 min (X), and the chromatogram is normally free of any other peaks. Calibration standards are prepared by spiking a stock solution of X into serum at 40, 120, and 160 ppm; these spiked calibrators are then extracted in the same manner as samples. A three-point calibration curve is run and if the regression is acceptable, this calibration curve is used for three months. With each batch of samples, a single injection of blank serum spiked to 160 ppm is made as a system suitability test; if this check sample assays at 160 ppm, the system is deemed stable and samples are run.

The method had been running acceptably until he ran out of the 1000 ppm stock of X used for spiking the check sample. When the new stock was prepared, the problem of a 400 ppm assay for the 160 ppm sample appeared.

Consider the Possibilities

In a case like this, I like to divide the case up into several possible problem areas, then see how many of these possibilities I can eliminate with the data at hand. This helps to focus my attention on the source of the problem so that it can be investigated further, if necessary, and corrected. We can broadly, and somewhat arbitrarily, divide the possible problem areas into chemistry, hardware, sample-related, and calibration. Let's look at each of these in more detail.

Chemistry

By chemistry, here I mean the chromatographically related chemical

John W. Dolan
LC Troubleshooting Editor

influences. These are the nature of the sample, the column, the mobile phase, and the column temperature. We can quickly eliminate these as the likely sources of the problem. If the column chemistry, mobile-phase chemistry, or column temperature had changed, we would expect a shift in retention for X and the IS, but this was not observed. The sample chemistry, or identity, is unlikely to have changed, because the check sample had no apparent retention problems in either the LC or GC assay.

Hardware

LC system hardware could malfunction in terms of flow rate, injection problems, or detection. The flow rate must be correct or the retention times would shift for both X and the IS. It is possible that the autosampler is not working properly, but this is unlikely to cause the noted problem, because any volume error in the autosampler would be compensated by the use of the IS. The purpose of the IS is to add it early in the sample preparation process so that any loss of sample volume or injection error would not matter, because it is the ratio of X/IS that is used in the calibration process, not the absolute response of either compound.

Problems related to the detector are a possible source of error, and should be checked. Two obvious possibilities are that the wrong wavelength was selected or that there is something wrong with the detector lamp. The response of X and the IS would be expected to change if the detection wavelength was changed, and a change in the relative response of X and the IS would be likely. This would generate a different X/IS ratio for a given concentration, which in turn would change the assay value for X in the check sample. A change in lamp energy as the detector lamp aged could also cause a change in response, and although I would expect that such intensity would affect X and the IS similarly, that is not a certainty. The proper wavelength should be verified and the lamp energy should be compared to normal values to determine if either of these items could be the problem source.

Sample-Related Problems

We know that the identity of the sample is correct and that the standard was made at the proper concentration because the sample assayed at 160 ppm by GC. The reader did not state if a new batch of IS stock was made at the same time, but if we consider the method, either the new batch of IS was made correctly or the old one was still good and was used. The check sample is made by spiking serum, and serum would never be injected directly, so it follows that the check sample was spiked with IS and extracted in the normal manner. One of the reasons for adding IS is to account for the inevitable changes in sample volume that take place during sample preparation.

Let's review the sample cleanup procedure: 300 µL of serum is combined with 50 µL of IS and 200 µL of dilute hydrochloric acid (550 µL total), centrifuged and extracted with 600 µL of dichloromethane. All of X and the IS should transfer into the dichloromethane, so the concentration of X and IS is 550/600 of its concentration in the original diluted serum. Next, 400 µL of the dichloromethane is removed, evaporated to dryness, and reconstituted in 50 µL of methanol. This concentrates the dichloromethane extract by 400/50 or eightfold. With the extraction, evaporation, and reconstitution steps, there will be inevitable volumetric errors introduced, which is why the IS is added — the same losses of X and IS should occur, so the X/IS ratio should stay constant. All this leads me to conclude that the GC method would be very unlikely to give an assay value of 160 ppm of X by an external standard method, even if the results were adjusted for the theoretical changes in concentration. Instead, I conclude that the IS method was used for GC, as well, and because the assay was as expected, it tells me that the check sample was made correctly, even though it doesn't assay properly by the LC method. The bottom line here is that it is unlikely that the current problem lies with the sample or sample preparation.

Calibration

At this point we've eliminated chemistry problems, hardware problems

(assuming the detector wavelength is set correctly and the detector lamp is in acceptable condition), and sample-related problems. This leaves calibration problems as the most likely problem source (assuming that we haven't overlooked something else obvious, which is always a possibility).

My initial interaction with the reader simply indicated that the check sample did not assay correctly by LC, but gave the expected answer by GC. When I requested more information about the method, I learned of the practice of calibrating every three months and using the system suitability check sample to verify that the method was working properly. Although the rules are a bit different in the clinical laboratory industry, this goes strongly against the analysis of the same drugs in serum or plasma to support drug development in the pharmaceutical industry. The latter techniques fall under guidelines from the United States Food and Drug Administration (FDA). The FDA's "Guidance for Industry: Bioanalytical Method Validation" (1) discusses validation of methods for the analysis of small molecular weight drugs in plasma and other tissues (generally called "bioanalytical" methods, as opposed to methods for the analysis of biological compounds). In this document in the section titled "Application Of Validated Method To Routine Drug Analysis" (pp. 13–14), it is stated:

> A calibration curve should be generated for each analyte to assay samples in each analytical run and should be used to calculate the concentration of the analyte in the unknown samples in the run . . . . The calibration (standard) curve should cover the expected unknown sample concentration range in addition to a calibrator sample at LLOQ [lower limit of quantification].

It goes on to say:

> Once the analytical method has been validated for routine use, its accuracy and precision should be monitored regularly to ensure that the method continues to perform satisfactorily. To achieve this objective, a number of QC [quality control] samples prepared separately should be analyzed with processed test samples at intervals based on the total number of samples. . . . The QC samples in duplicate at three concentrations . . .

Additionally, it is noted:

A matrix-based standard curve should consist of a minimum of six standard points, excluding blanks (either single or replicate), covering the entire range.

This says that the calibration curve should be run with each batch of samples, not once every three months. The calibration curve should cover the expected sample concentration range, and include the LLOQ. Furthermore, QC samples should be run at three concentrations that fall within the range of sample concentrations. These guidelines also make good sense from an analytical chemistry standpoint.

There are just too many potential problems that can occur that might cause the calibration curve to be different on different days. I have been involved with research and development (R&D) studies where the reference standards were so rare and valuable that it was not possible to run them every day, but a surrogate standard was found to verify that the original calibration was still adequate.

That may seem to align with the current problem, but in fact the drug X and its IS are very common compounds that can be purchased in reference standard grade for reasonable prices, so it is hard to justify trimonthly calibration on economic grounds.

The fact that the check sample was formulated at 160 ppm and verified by GC underlines the probability that the source of the problem lies with the calibration curve. My best guess is that something in the LC system has drifted over time, most likely the detector response (or an improper wavelength setting), and has caused the current response to the X/IS ratio to be much larger than it was when the calibration curve was run originally.

What Now?

I recommend that the proper wavelength setting and detector lamp performance be verified before proceeding. After these are found to be satisfactory, I would generate a new calibration curve using freshly prepared standards of X spiked into blank serum and extracted normally. I believe that the check sample will now assay correctly, closing the loop on identifying the problem source.

Technically, the check sample has done exactly what it was intended for — it has alerted the operator to a problem with the assay before valuable patient samples were run. However, I would modify the method to comply more with the industry standard of the FDA guidelines (1). This would require running a calibration curve, containing samples with at least six concentrations, each day with each batch of samples run. In addition, a set of check samples, or QCs, should be prepared and included in each sample batch to show that during the analysis, the method gives the expected results for samples of known concentration. There will be some documentation required to make these changes, but the method reliability will be much improved and should
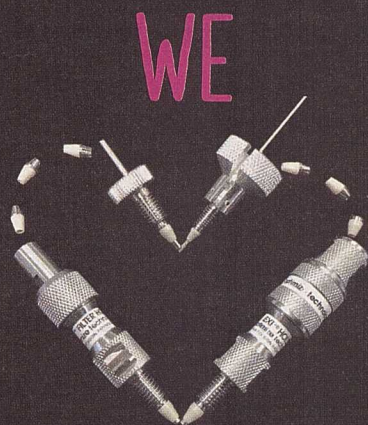
justify this extra work. The quality of the results produced should improve, as well. Finally, should the laboratory be audited by a regulatory agency, there will be much less likelihood of negative findings by the auditors.

In terms of day-to-day added work, there should be only a small impact on the total batch run time for a potentially large improvement in data quality. The calibration and check samples can be quickly spiked with known amounts of X and extracted with QC samples and samples to be analyzed. A total of six calibrators and six QC samples (duplicates at three concentrations) would add 12 samples to the day's run. At a 12-min retention time for X, this would increase the run time for the batch by about 2.5 h. It may be very easy to compensate for this increase in run time by increasing the flow rate; with an isocratic run, the separation should not be affected by the flow rate. The pressure would rise in proportion to the increase in flow rate, but it is fairly rare with conventional LC runs that pressure is a limit-

ing condition, so the added pressure is unlikely to be an issue.

Conclusions

We have used a specific example of a method problem to illustrate how to break down the problem into several potential problem sources. Most of these sources could be eliminated by careful consideration of the method and how the results deviated from the expected ones. This left us with two likely problem sources. First, a problem with the detector wavelength setting or detector lamp energy. These could be quickly checked by examining the instrument. The second potential problem source was that the instrument response to X or the IS had drifted between the time the original calibration curve was run and the problem was noted.

The recommended solution was to first check for detector problems, and second rerun the calibration curve. A more permanent fix to the problem would be to change the method to comply better with current FDA guide-

lines and general analytical chemistry practices of running calibrators contemporaneously with samples.

References

(1) United States Food and Drug Administration, *Guidance for Industry: Bioanalytical Method Validation* (FDA, Rockville, Maryland, 2001).

John W. Dolan
"LC Troubleshooting" Editor John Dolan has been writing "LC Troubleshooting" for LCGC for more than 30 years. One of the industry's most respected professionals, John is currently the Vice President of and a principal instructor for LC Resources in Lafayette, California. He is also a member of LCGC's editorial advisory board. Direct correspondence about this column via e-mail to John.Dolan@LCResources.com

For more information on this topic, please visit www.chromatographyonline.com/column-lc-troubleshooting