Application Note

Metabolomics/
Clinical Research

Agilent
Trusted Answers

Long-Term Robustness of Agilent LC/Q-TOF Systems for Untargeted Lipidomics

In large plasma cohort studies for clinical research

Authors

Tomas Cajka, Brian DeFelice, Gert Wohlgemuth, Tong Shen, and Oliver Fiehn
University of California, Davis Genome Center

Abstract

During large-scale lipidomics studies, it is necessary to minimize variability unrelated to the biological question under study. Robust instruments that provide stable precision and mass accuracy are essential to the quality of the untargeted lipidomics workflows used to inform large-scale studies. For example, precise measurements increase the statistical power of a study by reducing instrument error. This application note benchmarks the precision of the relative quantitative measurements and mass accuracy obtained for large-scale untargeted lipidomic analyses of human plasma samples using Agilent quadrupole time-of-flight liquid chromatography/mass spectrometry (LC/Q-TOF) systems. More than 14,000 plasma samples were analyzed over 26 consecutive months on 241 days of measurement. Lipidomic analyses were performed in both positive electrospray ionization (ESI) and negative ESI modes using C18-based reversed-phase chromatography and 15-minute data acquisition times. Quality control (QC) measures provided high-quality data across the sample set, including method blank controls, pool controls, and external community plasma controls (NIST SRM 1950). After Systematic Error Removal by Random Forest (SERRF) normalization on pool QCs and external test NIST QCs, the data quality obtained (<2% median RSD in pool QC samples and 8.3% median RSD in NIST pool QC samples) demonstrated reliable analysis of thousands of real-world plasma samples for quantitative evaluations of lipophilic compounds relevant to the development of type 1 diabetes. Analyses of internal standards in the pool and NIST QC samples also provided excellent mass accuracy (much better than the 2 mDa needed to obtain elemental formulas for unknowns) that was sustained over the time frame of the study. The results showed that the Agilent LC/Q-TOF systems are applicable to high-confidence, long-term, large-scale untargeted analyses in human plasma cohort studies.

Introduction

The largest sample sets for metabolomics and lipidomics analyses now originate from clinical and epidemiological cohort studies. To obtain statistical confidence when carrying out such large-scale studies, variability unrelated to the biological question being investigated must be controlled and minimized. From sample collection, handling, and preparation, to data acquisition by the instrument system, there are many possible sources of undesirable analytical variability. Ideally, the data acquisition system used should represent the lowest source of variability. Analytical reproducibility and robust mass accuracy are essential to untargeted lipidomic workflows. As previously demonstrated for a targeted plasma lipidomics workflow[1], high analytical reproducibility ensures that the data acquired directly reflect biological variance in population cohorts. It also reduces the number of technical replicates that must be tested, increasing workflow efficiency. Accurate-mass data increase parent and fragment ion specificity and thus identification certainty for unknown compounds. Though running an instrument continuously is preferred to meet throughput goals and minimize batch effects, data quality can suffer when doing so for too long without performing preventative maintenance.

This application note evaluates the quantitative and mass accuracy stability and performance of Agilent LC/Q-TOF systems for large-scale untargeted lipidomic analyses. The evaluation involved the analysis of more than 14,000 plasma samples over 26 consecutive months on 241 days of measurement. The plasma samples were collected from children diagnosed with type 1 diabetes mellitus (T1DM) for ongoing research in The Environmental Determinants of Diabetes in the Young (TEDDY) study. The TEDDY study is a multiyear, multicenter collaborative effort to find the causes of T1DM[2], which is usually diagnosed in childhood as insulin-dependent diabetes. While there are known genetic risk factors for T1DM[3], the environmental determinants of diabetes in the young are not known.

LC/Q-TOF MS lipidomic analyses were performed in both positive ESI and negative ESI modes. The sheer number of samples necessitated multiple quality control measures to ensure high-quality measurements across the entire set of samples, including method blank controls, pool controls, and external community plasma controls (NIST SRM 1950). Analyses were performed with methods established at the University of California (UC) Davis West Coast Metabolomics Center, using C18-based reversed-phase chromatography and 15-minute data acquisition times, similar to the methodology published by the authors in a previous application note.[4] As the study was funded by the National Institutes of Health, all data were uploaded to the NIH Metabolomics Workbench repository for retrospective analyses.

Experimental

Consumables and supplies

LC/MS-grade solvents and mobile phase modifiers were obtained from Fisher Scientific, Waltham, MA (water, acetonitrile, and methanol) and Sigma-Aldrich/Fluka, St. Louis, MO (isopropanol, formic acid, ammonium formate, methyl *tert*-butyl ether, and toluene). Internal standards for lipidomics analyses used odd-chain and deuterated chemicals: LPE 17:1, LPC 17:0, PC 12:0/13:0, PE 17:0/17:0, PG 17:0/17:0, d_7-cholesterol, SM d18:1/17:0, Cer d18:1/17:0, sphingosine d17:1, MG 17:0, DG 12:0/12:0, DG 18:1/2:0, d_5-TG 17:0/17:1/17:0, and CE 22:1.

Sample preparation

For each sample, 10 µL of blood plasma were extracted with a biphasic solvent system of cold methanol, methyl *tert*-butyl ether (MTBE), and water containing internal standards, as described previously.[4] Dried lipid extracts were resuspended using a methanol:toluene (9:1, v:v) mixture (150 µL for positive ESI with the Agilent Jet Stream Technology ion source (AJS(+)) and 50 µL for negative ESI with the AJS(−)), vortexed for 10 seconds and centrifuged at 14,000 rpm for 2 minutes before LC/MS analysis.

LC/Q-TOF MS analyses

The chromatographic separation method most widely used in lipidomics is reversed-phase HPLC, which separates lipids based on their nonpolar fatty acyl moieties.[5] LC and MS instrument parameters are provided in Tables 1 and 2, respectively.

To obtain comprehensive results, lipids were analyzed in both positive and negative ESI modes. Negative ESI runs were performed on an Agilent 6550 iFunnel Q-TOF LC/MS (at 20,000 resolving power) and positive ESI data were acquired on an Agilent 6530 Q-TOF LC/MS (at 10,000 resolving power). Data-dependent MS/MS (MS2) acquisition was used for lipid identification, while MS1 data were used for quantification. The instruments were tuned using an Agilent tune mix. A reference solution (*m/z* 121.0509 and *m/z* 922.0098 for positive mode, and *m/z* 119.036 and *m/z* 980.0163 for negative mode) was used to correct small mass drifts during data acquisition. The Q-TOF parameters are listed in Table 2.

Table 1. LC instrument parameters.

Parameter	Value
LC	Agilent 1290 Infinity LC system with a pump (G4220A), column oven (G1316C), and autosampler (G4226A)
Analytical Column	100 × 2.1 mm, 1.7 µm C18 column
Column Temperature	65 °C
Sampler Temperature	4 °C
Injection Volume	1.67 µL (positive mode) 5 µL (negative mode)
Mobile Phase A, AJS(+)	60:40 (v:v) acetonitrile:water with 10 mM ammonium formate and 0.1% formic acid
Mobile Phase B, AJS(+)	90:10 (v:v) Isopropanol:acetonitrile with 10 mM ammonium formate and formic acid (0.1%)
Mobile Phase A, AJS(–)	60:40 (v:v) Acetonitrile:water with 10 mM ammonium acetate
Mobile Phase B, AJS(–)	90:10 (v:v) Isopropanol:acetonitrile with 10 mM ammonium acetate
Flow Rate	0.6 mL/min
Gradient	Time (min) %A %B 0 85 15 2 70 30 2.5 52 48 11 18 82 11.5 1 99 12 1 99 12.1 85 15 15 85 15
Needle Wash	10 seconds flush port (isopropanol)
Injector Cleaning	Time 1: 0.1 minutes (bypass) Time 2: 11.6 minutes (mainpass/bypass) Time 3: 13.0 minutes (mainpass/bypass)

Table 2. Q-TOF instrument parameters.

Parameter	Value	Value
MS	Agilent 6550 iFunnel Q-TOF LC/MS with Agilent Jet Stream Technology ion source	Agilent 6530 Q-TOF LC/MS with Agilent Jet Stream Technology ion source
Source Parameters		
Gas (Nitrogen) Temperature	200 °C	325 °C
Gas Flow	14 L/min	8 L/min
Nebulizer	35 psi	35 psi
Sheath Gas (Nitrogen) Temperature	350 °C	350 °C
Sheath Gas Flow	11 L/min	11 L/min
Capillary Voltage	3.5 kV	3.5 kV
Nozzle Voltage	1 kV	1 kV
Mass Spectrometer Parameters		
ESI Polarity	Negative	Positive
Extended Dynamic Range	2 GHz	2 GHz
MS1 and MS2 Acquisition Speed	2 spectra/s	2 spectra/s
MS1 and MS2 Mass Range	60 to 1,700 *m/z*	60 to 1,700 *m/z*
Collision Energy	40 eV	25 eV

Data processing

Untargeted features were detected in MassHunter Qualitative analysis software and imported into Agilent Mass Profiler Professional (MPP) software for peak alignment and filtering to generate a peak list containing unique retention times and accurate masses. The LipidBlast library (MassBank.us) was used to annotate peaks with the MS2 data collected from the pooled samples. The target list and annotated identifications were processed in MassHunter Quantitative Analysis software to obtain peak heights and peak areas.

Sequence and maintenance schedule

Data were acquired over 26 consecutive months on 241 days of measurement. A total of 15,901 injections were performed on each Agilent LC/Q-TOF system, which consisted of 12,986 plasma samples, 1,382 pool quality control samples (QC), 1,382 method blanks, and 151 NIST (NIST SRM 1950) QC samples. Each set of 10 plasma samples was bracketed with a pooled QC sample and a method blank to control for cross-contamination and enable post-hoc data normalizations. Up to 85 injections were performed per day on each instrument. At 15-minute cycle times injection-to-injection, the sample analysis time accounted for almost 90% of the maximum run time possible per day, given the time needed for exchanging solvent bottles and vial trays.

Apart from weekends, the sample-analysis sequence was interrupted for more than a week 14 times within the study time frame for instrument maintenance, vacation, or to carry out other research projects that required timely completion. Such interruptions, especially for instrument maintenance, are unavoidable in large-scale, long-term projects. Drifts in instrument sensitivity were observed in response to deposition of nonvolatile materials on the ESI cone or ion optics. Maintenance operations were performed based on predefined quality control limits, using both internal standard intensity and peak shapes and the results from data analyses of pool QC samples. Except for sudden sensitivity drops or LC pressure issues, maintenance was scheduled after completion of sub-batches of 300 plasma samples and was typically performed in less than one week. Sensitivity was re-established after maintenance and instrument tuning, and sequences began anew.

Results and discussion

Precision with data normalization

Apart from compound identification, good precision of peak heights over time is the most important consideration in untargeted metabolomics and lipidomics. However, drifts and interruptions that lead to differences in relative peak intensities are expected. Precision is best assessed by analysis of coefficient of variance (CV), which is also called relative standard deviation (RSD). This study used median RSD over all detected metabolites, calculated from repeated analyses of the pooled QC samples to assess precision. While 30% RSD is typically acceptable in metabolomics and lipidomics studies[6,7], such deviation is wider than the actual biological differences in plasma. Therefore, optimized data normalization schemes were applied to account for both signal drifts and signal jumps. Different lipid classes showed different amounts of quantitation drifts that could not be normalized by simple batch corrections. Similarly, neither normalization to individual internal standards, nor correction to the sum of all internal standards, improved the precision of overall data.

Prior to this study, use of locally estimated scatterplot smoothing (LOESS)[8] had been recommended to correct for drifts in large studies, by following the drifts in peak intensities from QC sample to QC sample. While LOESS corrections led to a median precision of 12% RSD for the internal standards in the pool QCs, both identified and unknown lipids showed technical errors of >30% RSD that were unacceptable for the TEDDY children's cohort study. (Note: these were interday precision values obtained across the study timeframe. The intraday precision values were much lower.) To address this challenge, the SERRF[6] algorithm that summarizes drifts of peaks that show similar patterns was used instead. This machine-learning model improved precision to <2% median RSD in pool QC samples and 8.3% median RSD in NIST pool QC samples (Table 3). NIST pool QC samples served as independent test samples because they were not used in either the LOESS or SERRF model building, but likely overestimated the error rate because NIST plasma uses EDTA as anticoagulant, while both the TEDDY and pool QC samples used citrate.

Table 3. Interday precision (% RSD) after SERRF and LOESS normalization on pool QCs and external test NIST QCs across the entire approximately 2-year study. Values in red were considered unacceptable for the study.

	SERRF Normalization			LOESS Normalization		
	Internal Standards	Identified Lipids	Unknown Lipids	Internal Standards	Identified Lipids	Unknown Lipids
Pool QCs	1.8%	1.7%	1.8%	11.5 %	12.6 %	14.6 %
NIST QCs	8.9%	9.8%	6.3%	37.2 %	44.6%	30.8%

Overall, the data quality for this large-scale long-term cohort study showed that, with SERRF normalization, almost 13,000 real-world plasma samples were reliably analyzed for quantitative evaluations of lipophilic compounds that contribute to the development of type 1 diabetes. The difference in quantitative variance between the QC samples and the biological test samples in the TEDDY cohort is presented for two lipids in Figures 1A and 1B.

Mass accuracy stability

Apart from assessing plasma lipids in type 1 diabetes, the TEDDY consortium is interested in evaluating the impact of environmental exposure to chemicals, including the impact on currently unidentified MS peaks. For this purpose, using high-resolution exact-mass instruments obtained from Agilent LC/Q-TOF systems was necessary, because nominal mass instruments such as triple quadrupole mass spectrometers cannot derive elemental formulas for unknown compounds.

In this study, the stability of accurate-mass measurements was fundamentally important, because unknown compounds could be detected (and yield data-dependent MS2 fragmentations) at any time during the study's 26-month time frame. Analyses of the accurate masses of internal standards in the pool and NIST QC samples showed excellent mass accuracy over the duration of the study (Figure 2), demonstrating that Agilent LC/Q-TOF systems are well suited for long-term, large-scale untargeted analyses in human plasma cohort studies. Table 4 shows the average and median mass errors, in addition to the 95% confidence interval, for six example internal standards over all QC samples. With better than 2 mDa mass error, elemental formulas for unknown compounds can be calculated with high confidence.

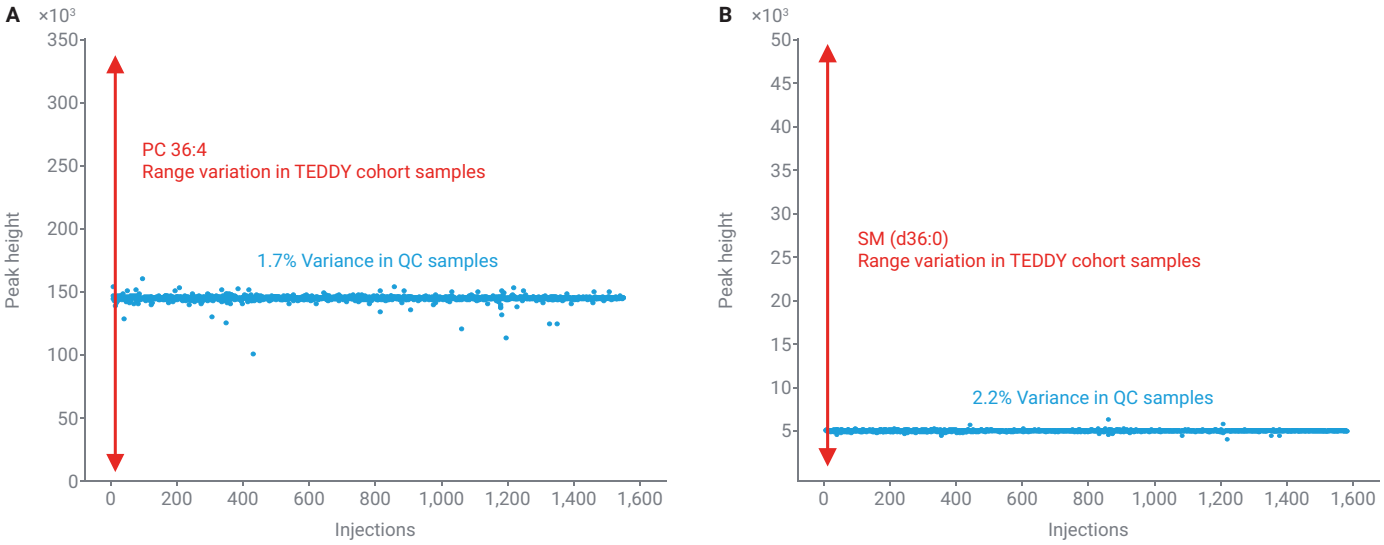

Figure 1. Examples of data variance in QC samples compared to the endogenous levels of plasma lipids in the TEDDY cohort. (A) phosphatidylcholine (36:4). (B) Sphingomyelin (d36:0). Data were for 1,533 pool and NIST quality control samples measured over 26 months.

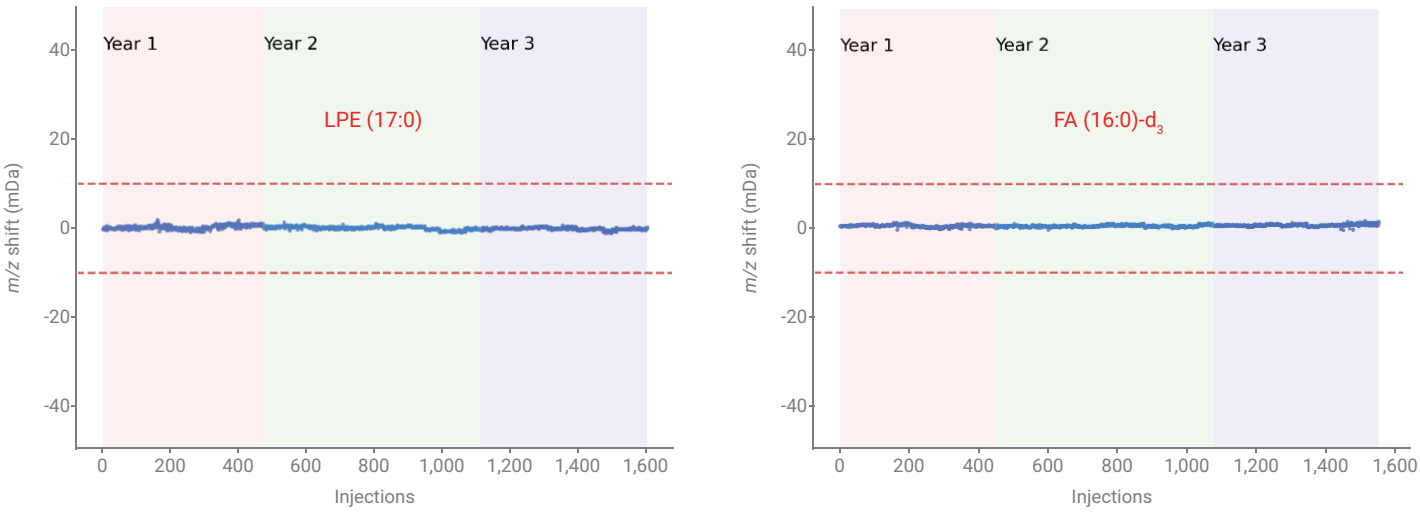

Figure 2. Example of mass accuracy measured by the 6550 IFunnel Q-TOF LC/MS in negative ESI mode for the QC samples lysophosphatidylethanolamine (LPE) 17:0 (A) and d_3-deuterated palmitic acid (B) compared to typically accepted thresholds for accurate masses at ±0.010 Da (±10 mDa, red dotted lines). Data were for 1,533 pool and NIST QC samples measured over 26 months.

Table 4. Mass errors for selected internal standards analyzed in QC samples over 26 months using Agilent LC/Q-TOF systems for negative and positive ESI measurements.

		Theoretical Mass	Avg Error (mDa)	Median Error (mDa)	SD (mDa)	95% Confidence (mDa)
FA 16_0 d_3	ESI neg	258.2518	0.51	0.50	0.25	1.0
CUDA ISTD	ESI pos	341.2799	0.36	0.30	0.27	0.9
LPE(17 1)	ESI neg	464.2783	0.31	0.26	0.26	0.8
LPC 17 0	ESI pos	510.3554	0.85	0.85	0.62	2.1
PE 17 0-17 0	ESI pos	720.5538	0.58	0.46	0.51	1.6
SM (d18 1-17 0)	ESI neg	761.5814	0.92	0.86	0.63	2.2

Conclusion

To carry out large-scale lipidomics studies with a high degree of confidence, it is necessary to minimize variability irrelevant to the biological question being investigated. Analytical variability represents one component of unwanted variability. Therefore, robust instrument precision and mass accuracy are essential for the untargeted LC/MS lipidomic workflows used to inform large-scale studies. High analytical reproducibility ensures that the data acquired directly reflect the biological variance in sample cohorts and increases workflow efficiency. Accurate-mass data enhance identification certainty for unknowns.

This application note benchmarked the precision and mass-accuracy performance obtained for large-scale untargeted lipidomic analyses of human plasma samples using Agilent LC/Q-TOF systems. The evaluation involved analyses of more than 14,000 plasma samples over 26 consecutive months on 241 days of measurement, which had been collected from children diagnosed with T1DM as part of the TEDDY study. LC/Q-TOF MS lipidomic analyses were performed in both positive and negative ESI modes using C18-based reversed-phase chromatography and 15-minute data acquisition times. Quality control measures provided high-quality data across the entire set of samples, including method blank controls, pool controls, and external community plasma controls (NIST SRM 1950).

The LC/Q-TOF MS data quality (%RSD) obtained showed that, with SERRF normalization on pooled QCs and external test NIST QCs, thousands of real-world plasma samples were reliably analyzed for quantitative evaluation of lipophilic compounds relevant to the development of type 1 diabetes. Analyses of internal standards in the pool and NIST QC samples yielded excellent mass accuracy that was sustained over the duration of the study, demonstrating that the Agilent LC/Q-TOF systems evaluated are applicable to high-confidence, long-term, large-scale untargeted analyses in human plasma cohort studies.

References

1. Huynh, K. *et al.* A Comprehensive, Curated, High-Throughput Method for the Detailed Analysis of the Plasma Lipidome. *Agilent Technologies application note*, publication number 5994-3747EN, **2021**.
2. TEDDY. The Environmental Determinants of Diabetes in the Young. https://teddy.epi.usf.edu (accessed May 19, 2023).
3. Steck, A. K.; Rewers, M. J. Genetics of Type 1 Diabetes. *Clin Chem.* **2011** Feb, *57(2)*, 176–85. doi: 10.1373/clinchem.2010.148221. Epub 2011 Jan 4. PMID: 21205883; PMCID: PMC4874193.
4. Cajka, T.; Fiehn, O. LC/MS Method for Comprehensive Analysis of Plasma Lipids. *Agilent Technologies application note*, publication number 5991-9280EN, **2019**.
5. Cajka T.; Fiehn, O. Comprehensive Analysis of Lipids in Biological Systems by Liquid Chromatography-Mass Spectrometry. *Trends Anal. Chem*. **2014**, *61*, 192–206
6. Fan, S. *et al.* Systematic Error Removal Using Random Forest for Normalizing Large-Scale Untargeted Lipidomics Data. *Anal. Chem.* **2019**, *91(5)*, 3590–6.
7. Dunn, W. *et al.* Procedures for Large-Scale Metabolic Profiling of Serum and Plasma Using Gas Chromatography and Liquid Chromatography Coupled to Mass Spectrometry. *Nat. Protoc.* **2011**, 6, 1060–1083. https://doi.org/10.1038/nprot.2011.335
8. Wang, S. Y.; Kuo, C. H.; Tseng, Y. J. Batch Normalizer: a Fast Total Abundance Regression Calibration Method to Simultaneously Adjust Batch and Injection Order Effects in Liquid Chromatography/Time-of-Flight Mass Spectrometry-Based Metabolomics Data and Comparison with Current Calibration Methods. *Anal. Chem.* **2013**, *85(2)*, 1037–46.

www.agilent.com

RA45209.6793634259

Printed in the USA, October 24, 2023
5994-6619EN

Agilent
Trusted Answers