Validation of a Direct Analysis in Real Time Mass Spectrometry (DART-MS) Method for the Quantitation of Six Carbon Sugar in Saccharification Matrix

Daudi Saang'onyo,[a,b‡] Gary Selby[b] and Darrin L. Smith*[a,b]

Eastern Kentucky University, [a]Department of Chemistry and [b]Center for Renewable and Alternative Fuel Technologies (CRAFT), 521 Lancaster Ave., Richmond, KY 40475, USA. Fax: 859-622-8197; Tel: 859-622-8380;

E-mail: daudi_saangonyo@mymail.eku.edu / gary.selby@eku.edu / darrin.smith@eku.edu

Electronic Supplemental Information (ESI)

Limit of Detection and Quantitation Data

The limit of detection deals with a peak three times the signal-to-noise ratio (S/N), and if the blank is taken as the y-intercept, Equation S1 can be used in the calibration curve in determining the LOD. Substituting the equation of the linear curve into Equation 1 gives Equation S2, which can be used to solve for x to give the LOD, expressed as x_{LOD} in Equation S3.[25]

$$y = 3s + b \qquad \text{Equation S1}$$

$$3s + b = mx + b \qquad \text{Equation S2}$$

$$x_{LOD} = \frac{3s}{m} \qquad \text{Equation S3}$$

In Equation S1, *b* is the y-intercept, and *s* is the standard deviation of the lowest concentration on the calibration curve. The LOD for glucose standards can then be estimated using the equations provided above. Using Equation S3, the standard deviation of the signal for the lowest concentration (0.00396) and the slope from the calibration curve (2035.2).

Concentration (M)	Average Peak Area Ratio (PAR)	PAR Standard Deviation
3.00×10^{-3}	6.17329	0.13575
2.00×10^{-3}	4.20513	0.01533
1.00×10^{-3}	2.14244	0.10152
6.00×10^{-4}	1.49796	0.01856
1.00×10^{-4}	0.46486	0.19118
6.00×10^{-5}	0.15259	0.01214
4.00×10^{-5}	0.09267	0.01995
1.00×10^{-5}	0.04111	0.00396

Table S1. For Limit of Detection (LOD) measurements, peak area ratios of different concentrations of glucose standards spiked with 4.00×10^{-4} M of internal standard (deuterated glucose) were generated along with the standard deviation for each meansurement (n=3).

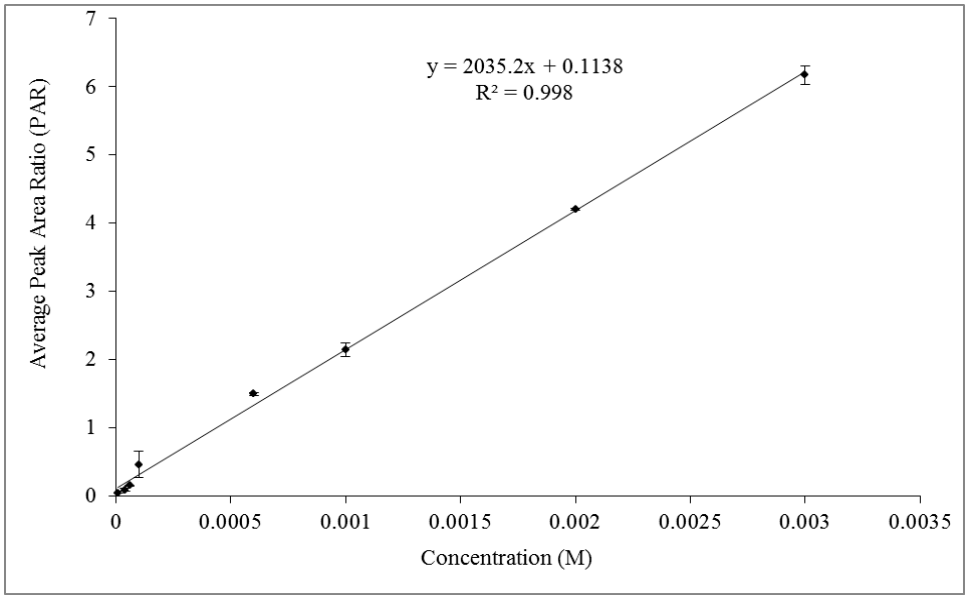

Figure S1. For Limit of Detection (LOD) determination, calibration curve for a series glucose standards solution spiked with 4.00×10^{-4} M of deuterated glucose (internal standard), data taken from Table S1. Each point represents an average (n = 3) peak area ratio with associated standard deviation.

An experimental confirmation was done to determine whether the calculated LOD is close to the detector signal produced at the estimated concentration. To fine-tune LOD determination, replicates of glucose standards whose concentrations were lower than the low-concentration sample in the calibration curve were prepared. This was done to get their S/N and determine if

their signals were less than three times greater than the noise. The S/N produced by the lowest concentration data point previously used in the calibration curve (1.00×10^{-5} M) was always above 5 (data not shown). The other glucose standards below that point (3.50×10^{-6}, 4.00×10^{-6}, and 5.00×10^{-6} M solutions of glucose standards), which were analyzed together with the lowest concentration value in the calibration curve to determine their S/N. It was observed that analysis of glucose concentrations near (or below) the calculated LOD gave an experimental signal with at least an S/N of 3 or less and in all the subsequent experiments performed (data not shown). For the LOQ, the value of ten replaced three in Equations S1–3. In most case cases, the LOQ is only estimated by observing the signal peaks at a concentration in which the S/N is at least 10. The S/N produced by 1.00×10^{-5} M (data not shown) was typically in the range of 5 and 10 and was estimated to be the LOQ. However, the calculated LOQ was found when ten is substituted for three in Equation S3 (LOQ = 10s/m). Since the volume of the glucose standard spiked on the glass tips was only 1.0 µL, the precise concentration of the LOD can be determined.

Calibration Curves Comparison: Statistical Analysis Calculations

The Student's *t*-test (commonly referred to as the *t*-test) was used for testing the difference between the available replicate measurements. The calculated *t* value ($t_{calculated}$) was then compared to the Student's *t* value (t_{table}) at the 95% confidence level for the corresponding number of degrees of freedom.

Using statistical analysis for data presented in Table 1 (in article), values of x and y that generate the best-fit (least squares) trend line are used to calculate the predicted values, $\hat{X}$ and $\hat{Y}$, respectively, from the straight lines. The $\hat{Y}$ value is solved from equations of the curves ($y = mx + b$, where m is the slope and b the y-intercept for each of the curves) by using the corresponding x values from each curve (matrix-free and matrix-diluted calibration curves).

After solving for $\hat{Y}$ values, the next step is to calculate the residual sum of squares (SS_{res}) for each curve using Equation S4.

$$SS_{res} = \sum_{i=1}^{n} (y_i - \hat{Y}_i)^2 \qquad \text{Equation S4}$$

The criterion used here is that of least squares, which considers the vertical deviation of each point from the line (i.e., the deviation we describe here as $(y_i - \hat{Y}_i)$, and defines the best-fit line as that which results in the smallest value for the sum of the squares of these deviations for all values of y_i and $\hat{Y}_i$. That is, $\sum_{i=1}^{n} (y_i - \hat{Y}_i)^2$ is to be a minimum, where n is the number of data points composing the sample. Once the SS_{res} has been determined, the mean square residual (MS_{res}) is computed using Equation S5 as a function of the residual degrees of freedom. MS_{res} defines the mean variance around the curves.

$$MS_{res} = \frac{SS_{res}}{n-2} \qquad \text{Equation S5}$$

where n is the number of data points composing the sample (at different concentrations), therefore, $n - 2$ is the residual degree of freedom defined by the difference of the total degrees of freedom and the regression degrees of freedom. From the mean square variance, the standard error of estimate, $S_{y \bullet x}$, (occasionally termed the "standard error of the regression") can be found according to Equation S6. The standard error of estimate is an overall indication of the accuracy with which the fitted regression function predicts the dependence of y on x.

$$S_{y \bullet x} = \sqrt{MS_{res}} \qquad \text{Equation S6}$$

The $S_{y \bullet x(p)}$ was used to calculate the pooled variance between the methanol/water standards curve and the 1% BES curve. The $S_{y \bullet x(p)}$ is a pooled standard deviation making use of both sets of data

(the matrix–free and matrix–diluted data). Once the $S_{y\bullet x(p)}$ has been determined, one can calculate $S_{b(p)}$.[25] The pooled error of the slopes was calculated using Equation S7.

$$s_{b(p)} = s_{y\bullet x(p)}\sqrt{\frac{1}{(\sum x^2)_1} + \frac{1}{(\sum x^2)_2}} \qquad \text{Equation S7}$$

where subscripts 1 and 2 refer to the two regression lines (the matrix–free and matrix–diluted lines, respectively) being compared, $s_{b(p)}$ is the variance for the slopes of each curve, and $s_{y\cdot x(p)}$ is the pooled standard error of the estimates, a measure of the uncertainty in the instrument response (y values on the calibration curves) and was calculated using Equation S8:[S1]

$$s_{y\bullet x(p)} = \sqrt{\frac{(n_1 - 2)(s^2_{y\bullet x(1)}) + (n_2 - 2)(s^2_{y\bullet x(2)})}{n_1 + n_2 - 4}} \qquad \text{Equation S8}$$

where $s^2_{y\cdot x(1)}$ and $s^2_{y\cdot x(2)}$ are the variances in the matrix-free and matrix-diluted datasets, respectively, and the factor $(n_1 + n_2 - 4)$ represents the pooled number of degrees of freedom. In Equation S7, $(\Sigma\, x^2)$ is defined as $\sum_{i=1}^{n}(x_i - \hat{X}_i)^2$. After computing the $S_{b(p)}$, the last step will be calculating the $t_{\text{calculated}}$ value (Equation S9) for the slopes to determine if the slopes are significantly different or not.

$$t_{calculated} = \frac{|b_1 - b_2|}{S_{b(p)}} \qquad \text{Equation S9}$$

where $|b_1 - b_2|$ is the absolute value of the difference of the slopes for matrix–free and matrix–diluted calibration curves. The $t_{\text{calculated}}$ from Equation S9 is compared with the t value in the Student's t table (t_{table}). If $t_{\text{calculated}}$ is greater than t_{table} at the 95% confidence level, the two slopes are considered to be significantly different.

Switchgrass Samples: DART-MS Analysis

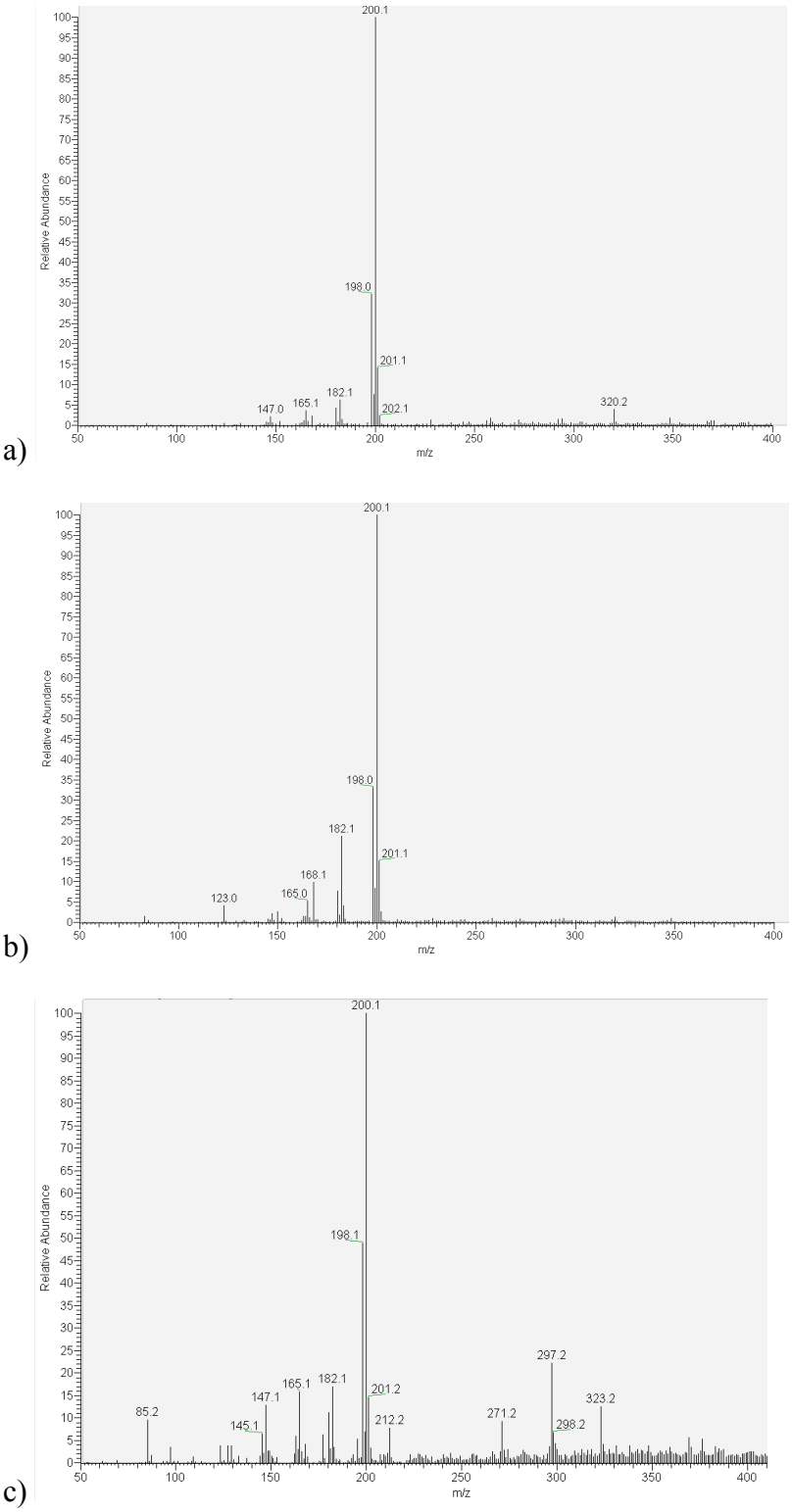

Figure S2. DART-MS analysis of switchgrass samples spiked with deuterated glucose (internal standard). Each switchgrass sample was pretreated using a) dilute acid, b) lime, and c) sulfuric acid before saccharification. Dominant ions for each spectrum are *m/z* 198, ammonium adduct of six carbon sugars and *m/z* 200, ammonium adduct for deuterated glucose, $[M\text{-}d_2 + NH_4]^+$. Tandem mass spectrometry for m/z 198 yielded fragmentation patterns similar to glucose (data not shown).

Recovery of Control Samples: Data and Statistical Analysis Calculations

Standards spiked into blank enzyme solution (matrix–diluted)						
	HQC (2.50×10^{-3} M)		MQC (1.50×10^{-3} M)		LQC (5.00×10^{-4} M)	
Concentration (M)	Average PAR[a]	Standard Deviation	Average PAR[a]	Standard Deviation	Average PAR[a]	Standard Deviation
3.00×10^{-3}	6.1383	0.1462	6.1955	0.0866	6.17928	0.1277
2.00×10^{-3}	4.0520	0.2351	4.1380	0.1344	4.1270	0.1627
1.00×10^{-3}	2.2260	0.0532	2.2362	0.0270	2.2557	0.0469
8.00×10^{-4}	1.8593	0.0670	1.8495	0.1624	1.8810	0.04192
4.00×10^{-4}	1.1024	0.0235	1.1314	0.0513	1.1620	0.0352
1.00×10^{-4}	0.2558	0.0190	0.2661	0.0630	0.2701	0.0446
QC Sample	5.1556	0.1014	3.1591	0.08138	1.2481	0.0263
Standards spiked into pure solvent (matrix–free)						
	HQC (2.50×10^{-3} M)		MQC (1.50×10^{-3} M)		LQC (5.00×10^{-4} M)	
Concentration (M)	Average PAR[a]	Standard Deviation	Average PAR[a]	Standard Deviation	Average PAR[a]	Standard Deviation
3.00×10^{-3}	6.2562	0.1777	6.3304	0.2632	6.2643	0.0586
2.00×10^{-3}	4.2548	0.0418	4.1991	0.0833	4.2010	0.0966
1.00×10^{-3}	2.1540	0.0676	2.1317	0.0830	2.1500	0.0402
8.00×10^{-4}	1.9421	0.1588	1.8356	0.0390	1.7992	0.0305
4.00×10^{-4}	1.1228	0.0305	1.1244	0.0150	1.1590	0.1189
1.00×10^{-4}	0.2218	0.0128	0.2173	0.0154	0.2190	0.0127
QC	5.0100	0.0945	3.0798	0.1466	1.2056	0.0276

Table S2. The average peak area ratios and the respective standard deviations from standards spiked into a pure and blank matrix solvents analyzed with QCs at three levels of concentration (low, mid, and high), data derived from Figures S3-8. [a]A blank was used to correct the PAR in each run.

Figures S3-5. A calibration curve obtained by analyzing blank matrix solutions spiked with glucose standards with high, mid-range, and low concentration quality controls (HQCs, MQCs, and LQCs, respectively). Each point is an average of three standard solutions with shown standard deviation.

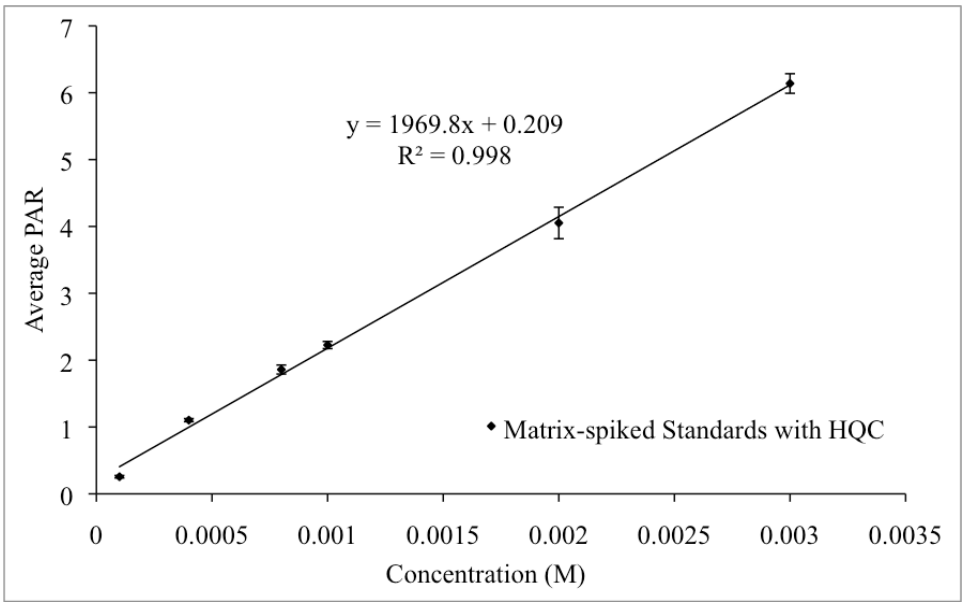

Figure S3. Generated calibration curve with matrix-diluted standards for analysis of HQC samples. Specific data to generate curve found in Table S2 and slope and intercept values are reported in Table 3.

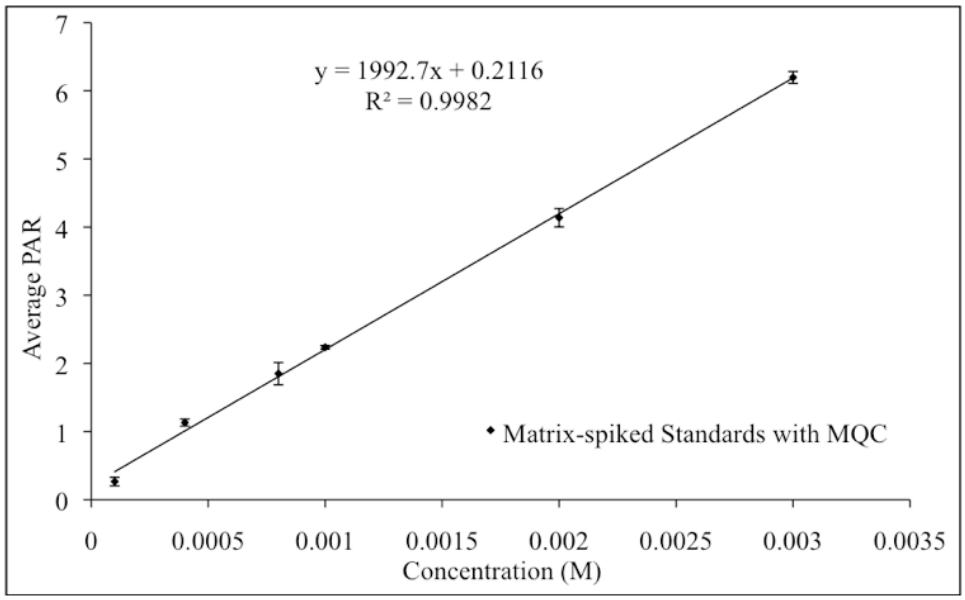

Figure S4. Generated calibration curve with matrix-diluted standards for analysis of MQC samples. Specific data to generate curve found in Table S2 and where slope and intercept values are reported in Table 3.

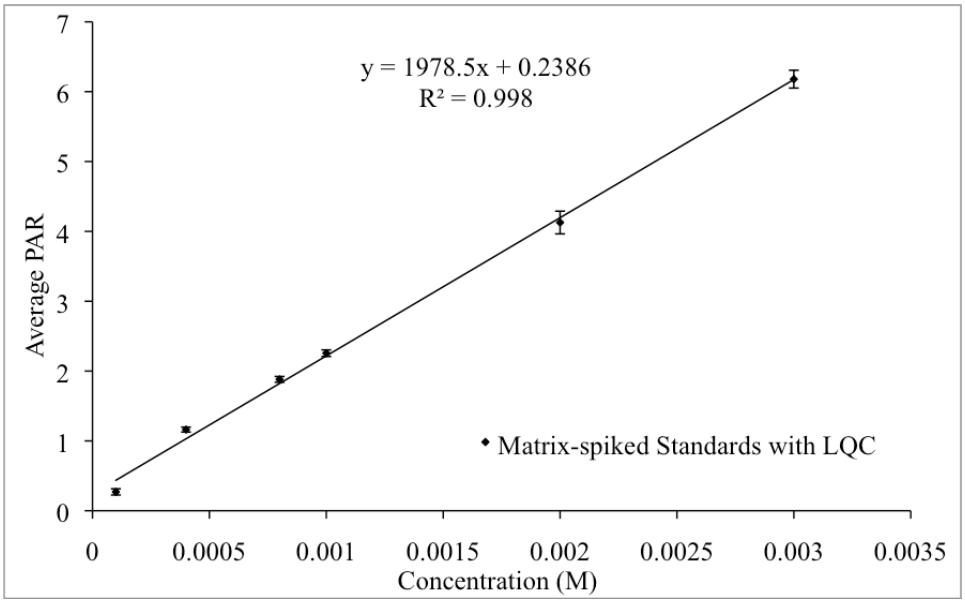

Figure S5. Generated calibration curve with matrix-diluted standards for analysis of LQC samples. Specific data to generate curve found in Table S2 and where slope and intercept values are reported in Table 3.

Figures S6-8. A calibration curve obtained by analyzing glucose standards prepared in pure solvents with high, mid-range, and low concentration quality controls (HQCs, MQCs, and LQCs, respectively). Each point is an average of three standard solutions with shown standard deviation.

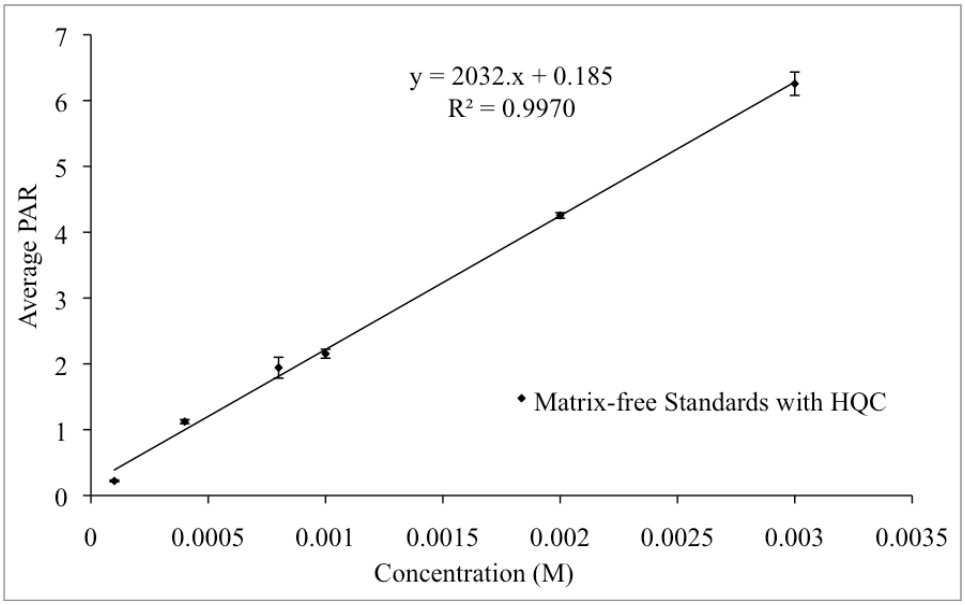

Figure S6. Generated calibration curve with matrix-free standards for analysis of HQC samples. Specific data to generate curve found in Table S2 and where slope and intercept values are reported in Table 3.

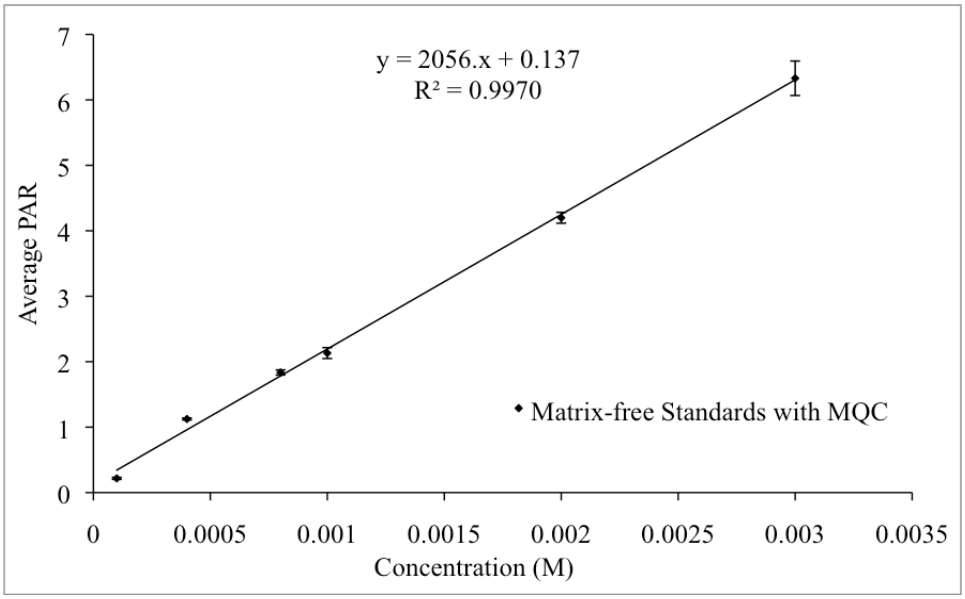

Figure S7. Generated calibration curve with matrix-free standards for analysis of MQC samples. Specific data to generate curve found in Table S2 and where slope and intercept values are reported in Table 3.

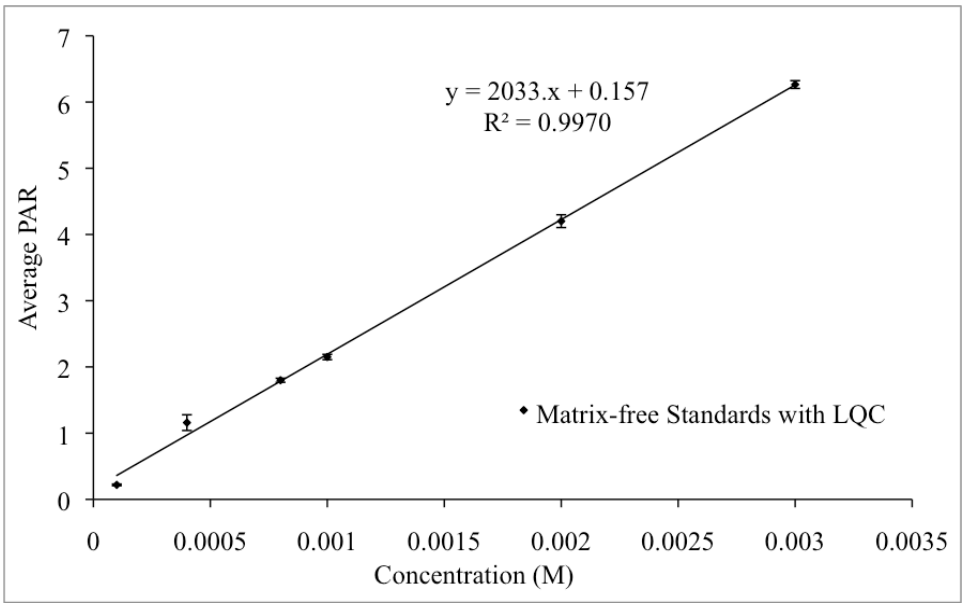

Figure S8. Generated calibration curve with matrix-free standards for analysis of LQC samples. Specific data to generate curve found in Table S2 and where slope and intercept values are reported in Table 3.

The PAR for each run of the QCs (whose concentrations were known) at the three concentration levels was obtained and used to compute the "calculated concentration" from the two sets of standards (matrix–free and matrix–diluted) and then generate an average calculated concentration at each QC level. The $\bar{x}_1$ and $\bar{x}_2$ were assigned as the average "calculated concentration" in the matrix–free set and matrix–diluted set, respectively. Each set of measurement has its own uncertainty and assume the population standard deviation (σ) for each set is essentially the same. Table 4 shows the data used for this analysis, where the label numbers 1, 2, and 3 indicate the three replicates for each QC. The s_1 and s_2 are assigned as the standard deviations for the matrix–free and matrix–diluted sets, respectively.

	Matrix-diluted Set		Matrix-free Set	
QC Samples	PAR	Calculated Concentration[b]	PAR	Calculated Concentration[b]
HQC1	5.1479	2.52×10^{-3}	4.9761	2.40×10^{-3}
HQC2	5.0582	2.45×10^{-3}	4.9362	2.32×10^{-3}
HQC3	5.2606	2.57×10^{-3}	5.1162	2.40×10^{-3}
	Mean ($\bar{x}_2$)	2.51×10^{-3}	*Mean* ($\bar{x}_1$)	2.37×10^{-3}
	SD[a] (s_2)	6.23×10^{-5}	*SD*[a] (s_1)	4.45×10^{-5}
MQC1	3.0846	1.47×10^{-3}	2.9543	1.35×10^{-3}
MQC2	3.2460	1.52×10^{-3}	3.2410	1.50×10^{-3}
MQC3	3.1467	1.45×10^{-3}	3.0442	1.45×10^{-3}
	Mean ($\bar{x}_2$)	1.48×10^{-3}	*Mean* ($\bar{x}_1$)	1.43×10^{-3}
	SD[a] (s_2)	3.27×10^{-5}	*SD*[a] (s_1)	7.89×10^{-5}
LQC1	1.2783	5.34×10^{-4}	1.2322	5.28×10^{-4}
LQC2	1.2359	4.93×10^{-4}	1.2077	5.07×10^{-4}
LQC3	1.2301	5.03×10^{-4}	1.1770	5.11×10^{-4}
	Mean ($\bar{x}_2$)	5.10×10^{-4}	*Mean* ($\bar{x}_1$)	5.15×10^{-4}
	SD[a] (s_2)	2.11×10^{-5}	*SD*[a] (s_1)	1.10×10^{-5}

Table S3. Replicate sets of measurements (n = 3 for each set) for the calculated concentration of the QCs at different levels of concentrations using the matrix–diluted and matrix-free standards. [a]SD is the standard deviation, [b]Concentration is in M

For the two sets of data consisting of n_1 and n_2 (where n = 3 for each set) measurements (with averages $\bar{x}_1$ and $\bar{x}_2$), we calculate the value of t with the formula

$$t_{calculated} = \frac{|\bar{x}_1 - \bar{x}_2|}{S_{pooled}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \qquad \text{Equation S7}$$

where $|\bar{x}_1 - \bar{x}_2|$ is absolute value of the difference of the means of the two sets and S_{pooled} is a pooled standard deviation making use of both sets of data:[25]

$$S_{pooled} = \sqrt{\frac{\sum_{set1}\left(x_i - \bar{x}_1\right)^2 + \sum_{set2}\left(x_j - \bar{x}_2\right)^2}{n_1 + n_2 - 2}} = \sqrt{\frac{s_1^2\left(n_1 - 1\right) + s_2^2\left(n_2 - 1\right)}{n_1 + n_2 - 2}} \qquad \text{Equation S8}$$

where $s_1$and s_2 are the standard deviations for the matrix–free and matrix–diluted standard sets. Using these equations, the $t_{calculated}$ values were computed from the data in Table S3 for each level of concentration of the QCs. These values are compared with the t values in the Student's t table (t_{table}) for $n_1 + n_2 - 2$ degrees of freedom as shown in Table S4.

QC Levels	S_{pooled}	$t_{calculated}$	t_{table} (95% confidence)	Do Measurements Agree?
HQC	2.71×10^{-5}	6.240	2.776	No
MQC	3.02×10^{-5}	1.923	2.776	Yes
LQC	8.40×10^{-6}	0.766	2.776	Yes

Table S4. The t test results for the statistical comparison of QCs calculated concentration with respect to the two sets of standards (matrix–free and matrix–diluted) for the three levels of QCs concentrations.

Supplemental Bibliographic Reference:

S1. D. C. Harris. *Quantitative Chemical Analysis*, 8th ed., W. H. Freeman & Co., New York, 2010.