Supporting Information for the Following Article in ACS Earth and Space Chemistry:

Comparison of Experimental *vs.* Theoretical Abundances of $^{13}CH_3D$ and $^{12}CH_2D_2$ for Isotopically Equilibrated Systems From 1-500°C

Daniel L. Eldridge*[1,2], Roman Korol[3], Max K. Lloyd[2], Andrew C. Turner[2], Michael A. Webb[4], Thomas F. Miller III[3], Daniel A. Stolper[1,2]

[1]*Energy Geosciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720 USA*
[2]*Department of Earth and Planetary Science, University of California, Berkeley, CA 94720 USA*
[3]*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125 USA*
[4]*Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544 USA*

*To whom correspondence should be addressed: danieleldridge@lbl.gov

Note: Supplementary tables (Table SI.1, SI.2, SI.3, SI.4, SI.5 and SI.6) are provided in an accompanying .xlsx spreadsheet.

SI.1 Fragment-correction scheme to determine δD, $\delta^{13}C$, $\Delta_{13_{CH_3D}}$, and $\Delta_{12_{CH_2D_2}}$

Here we describe how measured isotopologue ratios that include ion fragments in the denominator are corrected to obtain isotopologue ratios without the fragments and calculate δD, $\delta^{13}C$, $\Delta_{13_{CH_3D}}$, and $\Delta_{12_{CH_2D_2}}$ values.

The measured quantity that is used to derive the D/H ratio of a sample (and, thus, its δD value) is given by:

$$^{12}CH_3D(frag)R_{sample} = \frac{[^{12}CH_3D]}{[^{12}CH_4]+[^{12}CH_2D]} \quad \text{(SI.1)}$$

Here, we describe this isotopologue ratio as $^{12}CH_3D(frag)R_{sample}$ rather than $^{12}CH_3DR_{sample}$ (where $^{12}CH_3DR_{sample}$ = $[^{12}CH_3D]/[^{12}CH_4]$) to indicate that there is a fragment ('frag'; i.e., $^{12}CH_2D$) included in the denominator.

Following Stolper et al.[5] (see reference list in the main text), if we assume that the fragmentation ratio, F, which describes the rate of removing a hydrogen or deuterium from a methane isotopologue during ionization in the mass spectrometer, is (*i*) identical for all isotopologues ($^{12}CH_4$, $^{12}CH_3D$, and $^{13}CH_4$), is (*ii*) the same for H and D, and that (*iii*) isotopes are randomly distributed among all isotopologues, we can rewrite Eq. SI.1 as:

$$^{12}CH_3D(frag)R_{sample} = \frac{4\,{}^{D}R_{sample}}{1+3\times F\times {}^{D}R_{sample}} \quad \text{(SI.2)}$$

The three in the denominator is present as only ¾ of the time will $^{12}CH_3D$ fragment to $^{12}CH_2D$ in the ion source of the mass spectrometer. We have implicitly multiplied the ¾ by 4 to account for the conversion of $^{12}CH_3D$ abundances to a random isotopic distribution (i.e., $[^{12}CH_3D] = 4[^{12}C][H]^3[D]$ for a random isotopic distribution).

The quantity 'F' is measured operationally as the $[^{13}CH_3]/[^{13}CH_4]$ ratio as described in Section 2.4.4 of the main text. Rearrangement of Eq. SI.2 yields the desired quantity (^{D}R) in terms of the measured quantities (F and $^{12}CH_3D(frag)R$):

$$^{D}R_{sample} = \left(\frac{[D]}{[H]}\right)_{sample} = \frac{^{12}CH_3D(frag)R_{sample}}{4 - 3\times F\times {}^{12}CH_3D(frag)R_{sample}} \quad \text{(SI.3)}$$

An analogous treatment is applied to derive the ^{13}R of a sample for the computation of its $\delta^{13}C$ value. The measured quantity is:

$$^{13}CH_4(frag)R_{sample} = \frac{[^{13}CH_4]}{[^{12}CH_4] + [^{12}CH_2D] + [^{13}CH_3]} \quad \text{(SI.4)}$$

Applying the same assumptions as above we obtain:

$$^{13}CH_4(frag)R_{sample} = \frac{^{13}R_{sample}}{1 + F(^{13}R_{sample} + 3\,{}^{D}R_{sample})} \quad \text{(SI.5)}$$

Rearrangement yields:

$$^{13}R_{sample} = \left(\frac{[^{13}C]}{[^{12}C]}\right)_{sample} = \frac{^{13}CH_4(frag)R_{sample}\times(1 + 3\times F\times {}^{D}R_{sample})}{1 - F\times {}^{13}CH_4(frag)R_{sample}} \quad \text{(SI.6)}$$

Thus, in order to calculate a sample's ^{13}R value, we must first determine the ^{D}R value using Eq. SI.3.

From the $^{D}R_{sample}$ and $^{13}R_{sample}$ values, a sample's δD and $\delta^{13}C$ values can be computed. Implicit in the above derivations is the fact that we have removed the working gas that all sample measurements are made relative to by applying an analogous treatment to the working gas and computing a random distribution of its isotopologues based on the working gas's known D/H and $^{13}C/^{12}C$ ratios (as determined from the independent measurements made at UC Davis—see Section 2.1 and 3.1 in the main text).

The sample clumped isotopologue ratios ($^{13}CH_3D R_{sample} = [^{13}CH_3D]/[^{12}CH_4]$ and $^{12}CH_2D_2 R_{sample} = [^{12}CH_2D_2]/[^{12}CH_4]$) are derived similarly. Analogous to Eq. SI.4, the measured quantities are:

$$^{13}CH_3D(frag)R_{sample} = \frac{[^{13}CH_3D]}{[^{12}CH_4] + [^{12}CH_2D] + [^{13}CH_3]} \quad \text{(SI.7)}$$

$$^{12}\mathrm{CH_2D_2(frag)}R_{\mathrm{sample}} = \frac{[^{12}\mathrm{CH_2D_2}]}{[^{12}\mathrm{CH_4}]+[^{12}\mathrm{CH_2D}]+[^{13}\mathrm{CH_3}]} \quad \text{(SI.8)}$$

Applying the same assumptions as above yields:

$$^{13}\mathrm{CH_3D(frag)}R_{\mathrm{sample}} = \frac{^{13}\mathrm{CH_3D}R_{\mathrm{sample}}}{1+\mathrm{F}(^{13}R_{\mathrm{sample}}+3\,^{\mathrm{D}}R_{\mathrm{sample}})} \quad \text{(SI.9)}$$

$$^{12}\mathrm{CH_2D_2(frag)}R_{\mathrm{sample}} = \frac{^{12}\mathrm{CH_2D_2}R_{\mathrm{sample}}}{1+\mathrm{F}(^{13}R_{\mathrm{sample}}+3\,^{\mathrm{D}}R_{\mathrm{sample}})} \quad \text{(SI.10)}$$

Rearrangement yields:

$$^{13}\mathrm{CH_3D}R_{\mathrm{sample}} = {}^{13}\mathrm{CH_3D(frag)}R_{\mathrm{sample}} \times \left(1+\mathrm{F}\times\left(^{13}R_{\mathrm{sample}}+3\times\,^{\mathrm{D}}R_{\mathrm{sample}}\right)\right) \quad \text{(SI.11)}$$

$$^{12}\mathrm{CH_2D_2}R_{\mathrm{sample}} = {}^{12}\mathrm{CH_2D_2(frag)}R_{\mathrm{sample}} \times \left(1+\mathrm{F}\times\left(^{13}R_{\mathrm{sample}}+3\times\,^{\mathrm{D}}R_{\mathrm{sample}}\right)\right) \quad \text{(SI.12)}$$

Values of $\Delta_{^{13}\mathrm{CH_3D}}$ and $\Delta_{^{12}\mathrm{CH_2D_2}}$ relative to a random distribution for a given sample are then calculated as follows:

$$\Delta_{^{13}\mathrm{CH_3D}} = 1000\times\left(\frac{^{13}\mathrm{CH_3D}R_{\mathrm{sample}}}{4\times\,^{13}R_{\mathrm{sample}}\times\,^{\mathrm{D}}R_{\mathrm{sample}}}-1\right) \quad \text{(SI.13)}$$

$$\Delta_{^{12}\mathrm{CH_2D_2}} = 1000\times\left(\frac{^{12}\mathrm{CH_2D_2}R_{\mathrm{sample}}}{6\times\left(^{\mathrm{D}}R_{\mathrm{sample}}\right)^2}-1\right) \quad \text{(SI.14)}$$

Similar to δD and δ^{13}C values, implicit in the above derivations is the fact that we have removed the working gas that all sample measurements are made relative to by applying one of two treatments for the working gas composition:

(1) For compositions reported in the 'working gas reference frame', the $^{13}\mathrm{CH_3D(frag)}R_{\mathrm{sample}}$ and $^{12}\mathrm{CH_2D_2(frag)}R_{\mathrm{sample}}$values are calculated assuming a random distribution for the working gas $^{12}\mathrm{CH_3D}R_{\mathrm{wg}}$ and $^{12}\mathrm{CH_2D_2}R_{\mathrm{wg}}$ values, i.e.:

$$^{13}\mathrm{CH_3D(frag)}R_{\mathrm{wg}} = \frac{4\times\,^{\mathrm{D}}R_{\mathrm{wg}}\times\,^{13}R_{\mathrm{wg}}}{1+\mathrm{F}(^{13}R_{\mathrm{wg}}+3\,^{\mathrm{D}}R_{\mathrm{wg}})} \quad \text{(SI.15)}$$

$$^{12}\mathrm{CH_2D_2(frag)}R_{\mathrm{wg}} = \frac{6\times\left(^{\mathrm{D}}R_{\mathrm{wg}}\right)^2}{1+\mathrm{F}(^{13}R_{\mathrm{wg}}+3\,^{\mathrm{D}}R_{\mathrm{wg}})} \quad \text{(SI.16)}$$

As noted in the main text, compositions reported in the 'working gas reference frame' are denoted $\Delta_{^{13}\mathrm{CH_3D(wg)}}$ and $\Delta_{^{12}\mathrm{CH_2D_2(wg)}}$.

(2) For compositions reported in the absolute 'thermodynamic reference frame' in which Δ = 0‰ indicates the composition of a sample in internal isotopic equilibrium at infinite temperature, the ${}^{13}\mathrm{CH_3D(frag)}R_{\mathrm{sample}}$ and ${}^{12}\mathrm{CH_2D_2(frag)}R_{\mathrm{sample}}$ values are calculated using the determined ${}^{12}\mathrm{CH_3D}R_{\mathrm{wg}}$ and ${}^{12}\mathrm{CH_2D_2}R_{\mathrm{wg}}$ values for the working gas from the calibration of the current study (see Section 4.1 in the main text), i.e.:

$$ {}^{13}\mathrm{CH_3D(frag)}R_{\mathrm{wg}} = \frac{1.0025884\times\left(4\times {}^{\mathrm{D}}R_{\mathrm{wg}}\times {}^{13}R_{\mathrm{wg}}\right)}{1+\mathrm{F}\left({}^{13}R_{\mathrm{wg}}+3\,{}^{\mathrm{D}}R_{\mathrm{wg}}\right)} \quad \text{(SI.17)} $$

$$ {}^{12}\mathrm{CH_2D_2(frag)}R_{\mathrm{wg}} = \frac{1.0058600\times\left(6\times\left({}^{\mathrm{D}}R_{\mathrm{wg}}\right)^2\right)}{1+\mathrm{F}\left({}^{13}R_{\mathrm{wg}}+3\,{}^{\mathrm{D}}R_{\mathrm{wg}}\right)} \quad \text{(SI.18)} $$

SI.2 Peak Tailing Corrections on $^{12}CH_2D_2^+$ Intensity Using Water Peak Scans

We perform a correction to the measured $^{12}CH_2D_2^+$ intensity for tailing of signal from $^{13}CH_3D^+$ and $^{13}CH_5^+$. As discussed in the main text (Section 2.4.3), this follows the approach originally presented and worked out by Xie and co-workers[44] (see ref. list in main text).

The total background (b_{TOTAL}) for measured $^{12}CH_2D_2$ abundances is:

$$ b_{\mathrm{TOTAL}} = b_{\mathrm{off-peak}} + b_{\mathrm{tailing-}{}^{13}\mathrm{CH_5^+}} + b_{\mathrm{tailing-}{}^{13}\mathrm{CH_3D^+}} \quad \text{(SI.19)} $$

The $b_{\mathrm{off\text{-}peak}}$ represents the background of scattered ions present on all mass 18 species, which is measured during an off-peak background taken during the measurement (+0.1 Da off-peak; described in Section 2.4.3 of the main text). The $b_{\mathrm{tailing-}{}^{13}\mathrm{CH_5^+}}$ and the $b_{\mathrm{tailing-}{}^{13}\mathrm{CH_3D^+}}$ are the additional backgrounds caused by tailing of $^{13}CH_5^+$ and $^{13}CH_3D^+$, respectively.

Our methodology to determine $b_{\mathrm{tailing-}{}^{13}\mathrm{CH_5^+}}$ and $b_{\mathrm{tailing-}{}^{13}\mathrm{CH_3D^+}}$ is to perform scans of the resolved water peak that is located approximately 0.03 Da below the methane mass-18 ions ($^{13}CH_3D^+$, $^{13}CH_5^+$, $^{12}CH_2D_2^+$, $^{12}CH_4D^+$) (Fig. SI.1). In order to determine the amount of peak tailing from $^{13}CH_3D^+$ and $^{13}CH_5^+$, we use the $H_2{}^{16}O$ peak as a model peak shape. We use the water peak over those of the methane isotopologues because it is a more intense peak (~50,000 cps for water vs. <8000 cps $^{13}CH_5^+$ and $^{13}CH_3D^+$) and is fully resolved from other isobars, which allows for a more straightforward determination of peak tailing.

Scans of the water peak are performed under the same measurement conditions as $^{12}CH_2D_2$ including source pressure and tune settings (i.e., at the same mass resolving power) immediately prior to the $^{12}CH_2D_2$ measurement. From the water peak scan, the signal intensity is determined at two mass positions: the center of the water beam and the fixed mass differences (ΔM) from the center of the water peak that correspond to the mass differences between $^{13}CH_3D^+$ & $^{12}CH_2D_2^+$ ($\Delta M_{^{13}\mathrm{CH_3D^+}}$ = +0.00292 Da) and $^{13}CH_5^+$ & $^{12}CH_2D_2^+$ ($\Delta M_{^{13}\mathrm{CH_5^+}}$ = +0.00137 Da) (see Fig. SI.1). The measured intensities at the center of the water beam and at the fixed mass differences higher than the center position are first background-corrected based on the measured off-peak background

($b_{\text{off-peak}}$). The fixed-mass intensities are then ratioed against the intensity of the water peak at its peak center to obtain peak-tailing factors (*tf*), i.e.:

$$tailing\ factor = tf = \frac{i_{\text{tail}} - b_{\text{off-peak}}}{i_{\text{max}} - b_{\text{off-peak}}} \quad \text{(SI.20)}$$

Where i_{max} is the maximum intensity of the water peak at its center and i_{tail} is the intensity measured at +ΔM from the water peak center (ΔM = +0.00137 Da for $^{13}CH_5^+$-tailing, and +0.00292 Da for $^{13}CH_3D^+$-tailing; Figure SI.1). These peak-tailing factors are then applied to the measured maximum intensities (i.e. intensities at the peak center) for $^{13}CH_3D^+$ and $^{13}CH_5^+$ intensities (corrected for $b_{\text{off-peak}}$) for both the working gas and sample ($i_{^{13}CH_5^+} - b_{\text{off-peak}}$ and $i_{^{13}CH_3D^+} - b_{\text{off-peak}}$):

$$b_{\text{tailing}-^{13}CH_5^+} = \left(i_{^{13}CH_5^+} - b_{\text{off-peak}}\right) \times tf_{^{13}CH_5^+} \quad \text{(SI.21)}$$

$$b_{\text{tailing}-^{13}CH_3D^+} = \left(i_{^{13}CH_3D^+} - b_{\text{off-peak}}\right) \times tf_{^{13}CH_3D^+} \quad \text{(SI.22)}$$

The total background is then computed *via* Eq. SI.19 and subtracted from the measured $^{12}CH_2D_2^+$ intensities during data processing. Typical sizes of the tails are <0.4 cps for $b_{\text{tailing}-^{13}CH_5^+}$ and $b_{\text{tailing}-^{13}CH_3D^+}$ combined.

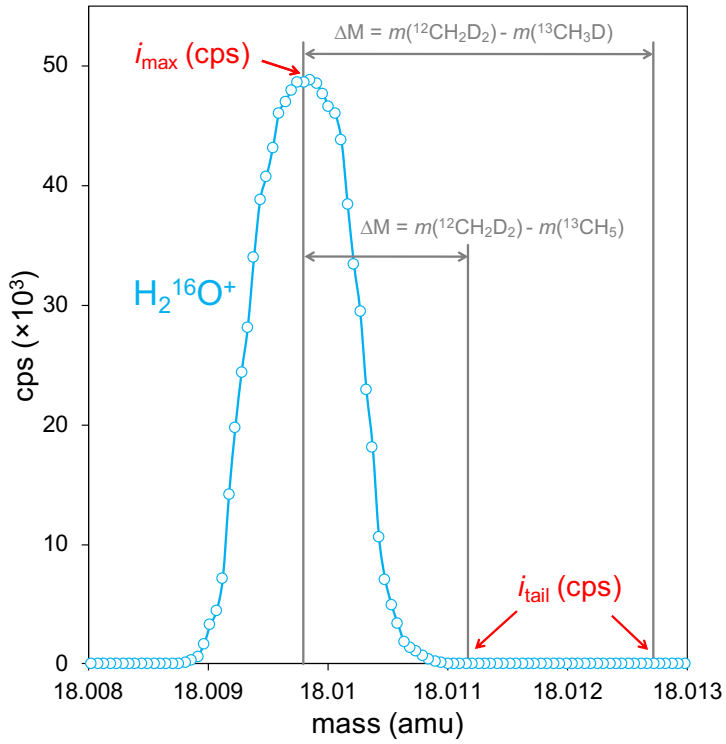

Figure SI.1: Water peak scan illustrating the procedure undertaken for tailing corrections on $\Delta_{^{12}CH_2D_2}$ measurements. Scan corresponds to the mass-18 scan in Fig. 1d in the main text.

The corrections to $\Delta_{12}{}_{CH_2D_2}$ due to this peak tailing are a function of the difference in δD between a sample and the working gas. Most of the experiments presented in the current study were performed with the working gas as the starting gas (Table SI.3). As a result, the consideration of peak tailing has little effect on the resulting $\Delta_{12}{}_{CH_2D_2}$ values. For example, the first experiment listed in Table SI.3 performed at 25˚C (δD 1‰ greater than the working gas) was determined to have a b_{TOTAL} = 0.491 cps (where $b_{\text{off-peak}}$ = 0.139 cps, hence peak tailing accounts for 0.352 cps of the total background on the $^{12}CH_2D_2$ intensity). The consideration of the tailing background in addition to the off-peak background ($b_{\text{off-peak}}$) results in a positive shift in $\Delta_{12}{}_{CH_2D_2}$ of ~0.2 ‰ relative to when only the off-peak background is considered, which is well within the internal precision of this measurement (±1.3 ‰, 1 s.e.). All equilibrated gas experiments conducted from 1-500˚C with the house methane exhibit shifts in $\Delta_{12}{}_{CH_2D_2}$ due to the peak tailing correction that are similarly small and are within the internal precision of the measurements. For heated gases (500˚C) exhibiting an approximate 200 ‰ range in δD, we see no significant dependence on values of $\Delta_{12}{}_{CH_2D_2}$ (Figure 2b of main text). From this exercise, we would expect significantly positive slopes in plots of $\Delta_{12}{}_{CH_2D_2}$ vs. δD if we were over-correcting our data due to tailing, and significantly negative slopes if we were under-correcting for tailing. Neither appears to be the case supporting the approach of our correction scheme.

Some samples included in the current study were measured prior to the practice of performing the detailed water scans, and therefore do not have directly constrained tailing backgrounds. These samples are noted in Tables SI.2 and SI.3. These samples also do not have direct constraints on the intensity of either $^{13}CH_3D^+$ or $^{13}CH_5^+$ on H4 CDD under the conditions of the measurements. Measurements performed without water scans are confined to some of the in-house reference standard measurements (especially the PlusD standard) and some of the heated gas experiments utilizing the Ni-based catalyst. For these samples, the background due to tailing was estimated from the average tailing backgrounds of subsequent sessions when tailing factors were constrained and observed to be stable, which comprise most (≥80%) of the data since water scans have been performed. These averages result in backgrounds due to tailing of: $b_{\text{tailing}-{}^{13}CH_5^+}$ = 0.30 cps and $b_{\text{tailing}-{}^{13}CH_3D^+}$ = 0.05 cps. We justify the application of these tailing backgrounds to measurements performed prior to performing daily water scans on the following grounds: (1) total tailing background corrections have thus far been observed to be stable over sessions in which the instrument was tuned to resolutions >28,000; (2) tune settings and mass resolving power of measurements prior to the practice of performing water scans were not significantly different than when water scans were performed; and (3) when the mean tailing backgrounds are applied to in house reference standards that did not have water scans they do not exhibit significantly different $\Delta_{12}{}_{CH_2D_2}$ values than samples that have water scans (Table SI.2). We also note that most of the equilibrated gases that do not have water scans are heated gas experiments performed with the Ni-based catalyst and prepared with the house methane. For such samples, corrections due to peak tailing have effects that are smaller than the internal precision of $\Delta_{12}{}_{CH_2D_2}$ measurements as discussed above.

SI.3 Data Filtering

We use the H4 Faraday cup to determine $^{12}\mathrm{CH_3D(frag)}R$ ratios and the H4 CDD secondary electron multiplier to determine the $^{12}\mathrm{CH_2D_2(frag)}R$ ratios. The H4 Faraday and the H4 CDD have the same sized exit slit width (0.04 mm) that is smaller than on the other detectors on the 253 Ultra (0.8 or 1.3 mm). This smaller exit slit makes isotopologue ratios obtained using the H4 detectors more sensitive to drift in the position of the ion beam relative to the mass scale over the course of a measurement as compared to detectors with wider exit slits. This is because the H4 cup exit slit width (0.04 mm) is similar in size to the width of the ion beam at medium- and high-resolution entrance slit configurations that are used for these measurements. As a result, drift in ion-peak position is a source of measurement imprecision in addition to other sources that include counting statistics (shot noise), scattered ion backgrounds, Johnson-Nyquist noise for Faraday cups, and dark noise for secondary electron multipliers.

We observe that the contribution of drift in ion-peak position to measurement precision scales with beam intensity. For example, the internal measurement precision for $^{12}\mathrm{CH_2D_2(frag)}R$, where the $^{12}CH_2D_2^+$ beam has intensities of 60-100 cps, corresponds to values expected from counting statistics (i.e., shot noise). In contrast, the $^{12}\mathrm{CH_3D(frag)}R$ measurement, where the $^{12}CH_3D^+$ beam has an intensity of ~1×10^6 cps, shows internal measurement precisions 3-4x greater than would be expected based on the combination of counting statistics and Johnson-Nyquist noise. We observe two kinds of drift in ion-peak position: (1) a higher-frequency oscillation of the beam location comprising small drifts in the mass scale on sub-second timescales; And (2) a lower-frequency oscillation of larger ion-peak position drift with a periodicity of tens of minutes due to sub-degree temperature fluctuations in the room. We apply two data filters to screen and account for both of these processes.

The higher-frequency (sub-second) ion-peak position oscillations cause the distributions of measured ion ratios (e.g., $^{12}\mathrm{CH_3D(frag)}R$) to be non-normally distributed. Specifically, distributions are skewed to values with lower ion ratios. This occurs because ion beams measured on the H4 cup have narrow or non-existent regions of peak flatness at the center of the peak. A flat peak occurs when the entire beam is in the cup without any tailing from other masses. When the beam is wider than the cup, the ion peak will appear rounded because the entire beam cannot enter the cup at any one mass position. For the H4 cup, we have observed that how rounded *vs.* flat the central area of the beam depends on the resolution of the instrument and the intensity of the beam. Drift in peak position off of the center position and onto a rounded flank of the peak lowers the measured isotope ratio regardless of the direction of the drift in peak position. If the beam is sufficiently intense, this isotope ratio will be lower at a resolvable level than the expected ranges of measured values based on consideration of shot noise, scattered ions, Johnson-Nyquist noise, and dark noise— we explain below how we determine the expected ranges. Thus, drift inherently skews measured isotope ratios to lower values relative to the true value, and more intense beams will appear more skewed. When the more intense $^{12}CH_3D^+$ beam is measured on the H4 cup using the medium-resolution slit with the HR+ aperture in, the peak shape is not perfectly flat. In contrast, the less-intense $^{12}CH_2D_2^+$ ion beam measured using the high-resolution slit with the standard aperture in shows less-rounded (i.e. flatter) peak tops. As a result, $^{12}CH_3D^+$ measurements are more sensitive to this high-frequency peak drift compared to measurements of $^{12}CH_2D_2^+$.

In order to account for this drift, we use an outlier test to filter data and exclude measured isotope ratios in which the ion beam has drifted sufficiently away from the center position to affect measurement quality. Outlier tests are commonly based on the number of standard deviations a measured sample is from the mean; a value of 3 standard deviations from the mean is a common rule of thumb to identify and exclude outliers. However, for our measurements on the H4 cup, both the measured mean values and standard deviations are affected by ion-position drift. This drift both skews measured mean values to lower values compared to the true value and increases the observed standard deviation *vs.* a situation without ion peak drift. As a result, using a conventional outlier test that excludes data points beyond a number of measured standard deviations from the measured mean will not be able to account for the peak-drift that is causing data to show larger scatter and lower mean value in the first place.

Instead, we base our outlier test on expected distributions of measured isotope ratios where counting statistics and additional error sources (e.g., Johnson-Nyquist noise for Faraday cups, dark noise for secondary electron multipliers, and scattered ions for both detector types) are the sole contributors to measurement uncertainty. In order to detect outliers on the timescale of the higher-frequency (sub-second) oscillations in ion-peak position, we filter data based on the distribution of sub-integrations from a single integration. As outlined in the main text, each measurement integration is purposefully discretized into sub-integrations with seventy-five 0.524 s sub-integrations for the ${}^{12}\mathrm{CH_3D(frag)}R$ measurement and sixty 1.048 s sub-integrations for the ${}^{12}\mathrm{CH_2D_2(frag)}R$ measurement.

As with typical outlier tests, we use the number of standard deviations, z (the so-called z score), that a sub-integration deviates from a mean value as a metric to decide whether a measured isotopic ratio is an outlier. We select the value of z to be used as a threshold value for our outlier test based on the number, n, of sub-integrations that comprise an integration. We now derive the relationship between z and n. For normally distributed data, the proportion of sub-integrations, P, that are expected to fall within z standard deviations of the mean of the integration is given by:

$$P = \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right). \tag{SI.23}$$

Because P is normalized such that it ranges from a value of 0 to 1, the proportion of sub-integrations that fall beyond z standard deviations of the mean is:

$$1 - P = 1 - \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right). \tag{S.24}$$

The value of 1–P represents the fraction of data from a normally distributed dataset that fall beyond z standard deviations from the mean. Thus, the value $1/(1-P)$ equals the number of observations, n, that would need to be made to typically observe a single sub-integration greater than or equal to z standard deviations from the mean. In other words, among a large series of normally distributed sub-integrations, there will be, on average, n consecutive sub-integrations between each sub-integration that falls at least z standard deviations from the mean. As a concrete example, if we measured 20 samples and 1 of those 20 samples is at the threshold value to be an outlier, then the value of 1-P is 0.95 and the value of z is 1.96. Based on this, we can rewrite Eq. SI.24 as

$$\frac{1}{n} = 1 - \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right). \quad \text{(SI.25)}$$

This can be rearranged as follows:

$$z = \sqrt{2} \times \mathrm{erf}^{-1}\left(1 - \frac{1}{n}\right). \quad \text{(SI.26)}$$

For the ${}^{12\mathrm{CH_3D(frag)}}R$ measurements, n is 75 (the number of sub-integrations) and thus z is 2.47. For the ${}^{12\mathrm{CH_2D_2(frag)}}R$ measurement, n is 60 and thus z is 2.39.

Eqs. SI.25 and SI.26 represent an idealized situation in which no drift in ion-peak position has occurred. As discussed above, drift in peak position will cause distributions of sub-integrations to be skewed to lower values. As a result, the mean value of such a population will be systematically lower than the true isotope ratio of the sample. To apply our filter, we need to exclude samples that are more than z standard deviations below the true mean value of the isotopic ratio being measured. To do this, we do not use the mean value of the measured distribution of isotope ratios due to its potential to be skewed to lower values *vs*. the true mean value. Instead, we use deviations of measured isotopic ratios values from the *maximum* measured ratios to establish if a sample is an outlier. We use maximum measured values as these are the least likely measured ratios to have been affected by drift in ion-peak position. This strategy is based on the assumption that for a normally distributed dataset, we are equally likely to have a sample measured that is z standard deviations above the mean as compared to below. Thus, by using maximum measured ratios, and looking at samples more than $2z$ standard deviations lower than it, we will generally filter out outliers that are at least z standard deviations below the true mean in the absence of ion-peak drift. We say generally above as inherent variability in measured ratios could cause exceptionally high or low maximum measured ratios to occur that are either greater than or less than z standard deviations from the mean. We discuss false-positive outlier identification rates below to address this.

Specifically, for each integration, we determine ratios for individual sub-integrations and exclude them (R_{out}) when their measured ratio is more than $2\times z$ expected standard deviations below the third-highest value of the distribution ($R_{3\mathrm{rdmax}}$):

$$\frac{R_{3\mathrm{rdmax}} - R_{out}}{\sigma_t} > 2 \times z. \quad \text{(SI.27)}$$

Here, σ_t is the expected standard deviation of sub-integrations based on counting statistics (σ_c) and other sources of measurement error unrelated to peak drift (σ_j, e.g., Johnson-Nyquist noise, dark noise and scattered ions). Let i be the intensity at the center position of the ion beam measured on H4 (either ${}^{12}CH_3D^+$ or ${}^{12}CH_2D_2^+$), i_b be the intensity of the mass 16 beam (>99% ${}^{12}CH_4^+$), and t_i be the duration of each sub-integration. The mass-16 beam is many orders of magnitude more abundant than these heavier isotopologues and a such we assume it contributes no error to the overall measurement. Thus, if shot noise were the only contributor to measurement uncertainty,

then the expected standard deviation of the ratio R (=i/i_b) of all sub-integrations in a single integration would be approximately:

$$\sigma_t = \sigma_c = \frac{\sqrt{i \times t_i}}{i_b}. \quad \text{(SI.28)}$$

However, detectors have additional sources of uncertainty. For measurements on a Faraday cup (as is the case for the $^{12}\mathrm{CH_3D(frag)}R$ measurement), the Johnson-Nyquist noise of the amplifier contributes to the expected standard deviation of the measurement. Johnson-Nyquist noise is specific to each amplifier and the duration of the sub-integration, but is generally important to consider for measurements on beams with intensities of 200 mV or less. The $^{12}CH_3D^+$ beam has an intensity of 100–160 mV in our measurement setup for $^{12}\mathrm{CH_3D(frag)}R$, so Johnson-Nyqvist noise contributes (see below). Two additional sources of error, dark noise and scatted ions, are detectable on the H4 CDD detector and typically contribute ±0.1 cps (1σ) of noise.

Regardless of the physical source of intrinsic detector uncertainty, our treatment is the same: during every analysis, we empirically determine the intrinsic noise level of the detector by calculating the standard deviation of sub-integrations measured during the off-peak background with gas flowing (see Section 2 of the main text, Methods). This empirical σ_j is agnostic of the source of noise, but markedly different on different detectors. For example, for the $^{12}\mathrm{CH_3D(frag)}R$ measurement, σ_j (mostly Johnson-Nyqvist noise) is typically 2000 cps, which is ~3X the shot noise ($\sigma_c \approx 700$ cps). For the $^{12}\mathrm{CH_2D_2(frag)}R$ measurement, σ_j is always less than 0.1 cps, and thus less than 2% of the shot noise. Regardless of the size of σ_j, the complete expression for the expected standard deviation of a population of sub-integrations is:

$$\sigma_t = \frac{\sqrt{i \times t_i + {\sigma_j}^2}}{i_b}. \quad \text{(SI.29)}$$

Our choice of applying the outlier test based on the third highest value observed (as opposed to the highest, second highest, or fourth highest, for example) is a compromise between our desire that the test be sufficiently sensitive to exclude times where ion-peak drift from the peak center affects data quality, but with a minimal rate of false positives. In practice, we find no change in results beyond stated error when using the first to sixth highest measured R value of a given set of sub-integrations. For example, for a δD measurement with a reported precision of ±0.15‰, using any of the six highest measured $^{12}\mathrm{CH_3D(frag)}R$ values for the outlier test on each integration changes the mean value and standard error of the measurement by less than 0.01‰. What does change, however, is the fraction of sub-integrations that are rejected as outliers. For example, for the $^{12}\mathrm{CH_3D(frag)}R$ measurements, using the third highest value, based on a Monte-Carlo simulation of 10^6 sets of 75 random samples from a normal distribution, this outlier test has a false-positive rate of 1 in 784 sub-integrations. For the $^{12}\mathrm{CH_3D(frag)}R$ measurement, this is equivalent to screening out data that are more than 3.0 standard deviations below the mean. For this measurement, $^{12}\mathrm{CH_3D(frag)}R$, this amounts to 8 false positives on average among the 6300 sub-integrations. As will be discussed below, the actual number of outliers identified is two orders of

magnitude greater. For the ${}^{12}\mathrm{CH_2D_2(frag)}R$ measurements, we are on average screening out data that are 2.95 standard deviations below the mean with a typical false positive rate of 1 in 635. For this measurement, ${}^{12}\mathrm{CH_2D_2(frag)}R$, this causes the identification of 36 false positives of the 22,680 sub-integrations.

In summary, combining eqns. SI.26, SI.27, and SI.29, the full outlier test for high-frequency ion-peak drift amounts to excluding all sub-integrations from each integration that meet the following condition:

$$R_{out} < R_{3^{rd}\max} - 2\sqrt{2} \times \mathrm{erf}^{-1}\left(1 - \tfrac{1}{n}\right) \times \frac{\sqrt{i \times t_i + {\sigma_j}^2}}{i_b}. \quad \text{(SI.30)}$$

We find for the ${}^{12}\mathrm{CH_3D(frag)}R$ measurement, on average 17% of sub-integrations are removed by this outlier test. Recall, that the rate of false positives expected is 0.13%. For a compilation of 80 δD measurements, the average measurement precision is improved from 0.16‰ to 0.12‰ when this filter is applied — i.e., a 25% improvement in precision. We note that this improved precision is still ~2 times higher than the shot noise + Johnson noise limit for the measurement; this filter is unable to completely eliminate the noise of ion-peak drift because much of it occurs within a few standard deviations of the true mean. In contrast, using a typical 3σ outlier test improves measurement precision from 0.158‰ to 0.156‰.

Importantly, measured mean δD values are not strongly affected by the filter— filtered data is, on average, 0.06‰ lower than unfiltered data and thus changed less than ±1 s.e. (0.12‰). Our filtering changes the mean measured δD values less than the standard deviation because δD values can be either increased or decreased due to variation in both sample and standard ${}^{12}\mathrm{CH_3D(frag)}R$ values. Additionally, external precision as determined from repeated measurements of in-house reference standards (± 0.15‰ in δD) is comparable to the internal precision (±0.12‰ in δD), indicating that the filter is not biasing our stated internal precision to too low a value.

For ${}^{12}\mathrm{CH_2D_2(frag)}R$ determinations, the average percentage of sub-integrations removed by this filter (0.16%) is not significantly higher than the expected false-positive rate. This indicates that the noise due to ion-peak drift on $^{12}CH_2D_2^+$ is controlled by the shot noise of the measurement. Thus, although we routinely monitor both unfiltered and filtered ${}^{12}\mathrm{CH_2D_2(frag)}R$ data, in practice we find insignificant differences in mean values and precisions when this filter is applied. Thus, the filter is not used to calculate final $\Delta_{12}{}_{\mathrm{CH_2D_2}}$ values. We note, however, that if a more intense ion beam is measured on the H4 CDD (for example, $^{13}CH_3D^+$), improvements in data precision can be observed when this filter is applied.

Although this filter improves measurement precision on high intensity beams (>10,000 cps beams), for the ${}^{12}\mathrm{CH_3D(frag)}R$ measurement there are also integrations where aberrant off-peak behavior is observed that require an additional correction. We observe this to occur when changes in laboratory temperature cause the ion beam to drift sufficiently off peak during a measurement

such that all sub-integrations of a given integration are potentially compromised. We can identify some measurement periods when this drift is occurring by observing when the fraction of an integration that fails the outlier test detailed in Eq. SI.30 is abnormally high (e.g. 40% relative to typical value of 17%). However, if this secular peak drift is sufficiently rapid such that the peak center is not accurately measured during the initial peak-center determination, the filter for higher-frequency peak-drift may not flag outliers as the measurement will start on a shoulder of the peak rather than the center. Instead, this behavior can be identified by comparing changes in measured peak-center positions of cycles of samples and standards. This comparison allows us to monitor for cycles where the peak center location moves faster than what is typically observed. Based on experience, we exclude all data from cycles where the maximum change in peak center between peak scans is equal to or greater than 40 μDa (equivalent to a secular drift rate of 0.45 μDa/s). In practice this approach eliminates, on average, 15% of all cycles, although this percentage changes on timescales of days to weeks depending on the temperature stability of the laboratory in a given session. We note that this cutoff is specific to the resolution settings and peak intensities (~1×10^6 cps) of this specific measurement design. Measurements at higher resolutions or lower peak intensities are less sensitive to the rate of secular peak drift. We have found that when peaks are less intense and measurements are shot noise-limited, as is the case for $^{12}\mathrm{CH_2D_2(frag)}R$ measurements, a cutoff at 0.45 μDa/s is excessive and unnecessarily excludes acceptable data. Therefore, this final data filtration is not used for $\Delta_{^{12}\mathrm{CH_2D_2}}$ measurements.

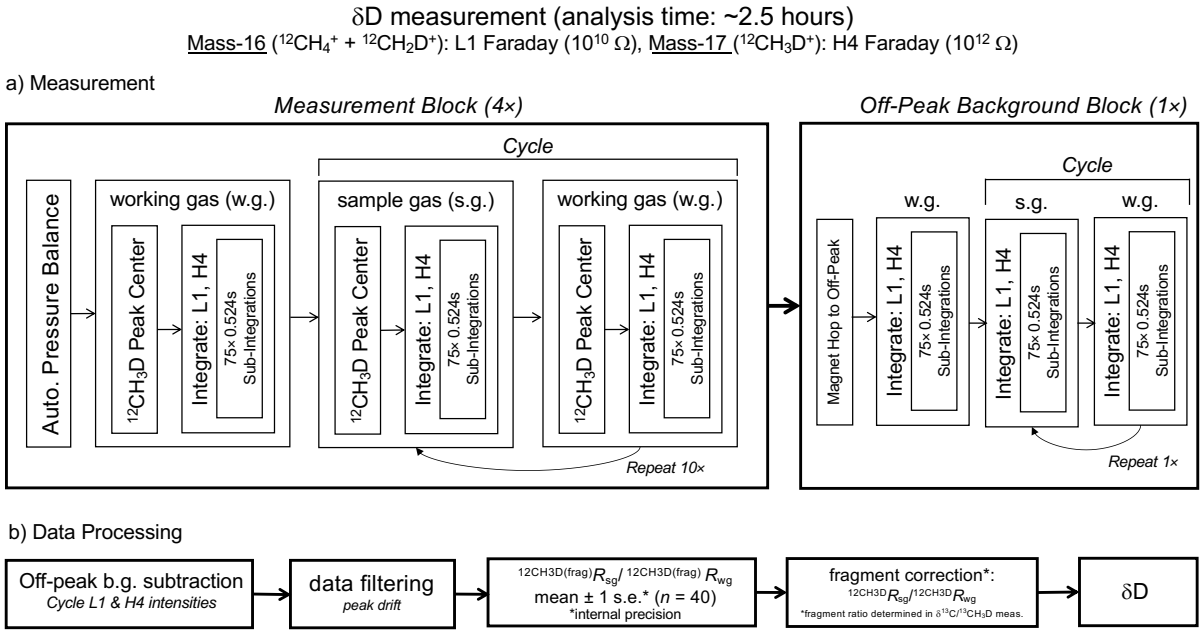

Figure SI.2(a-b): δD flow-chart diagram illustrating measurement and data processing details. See main text (Section 2.4.1) for further details.

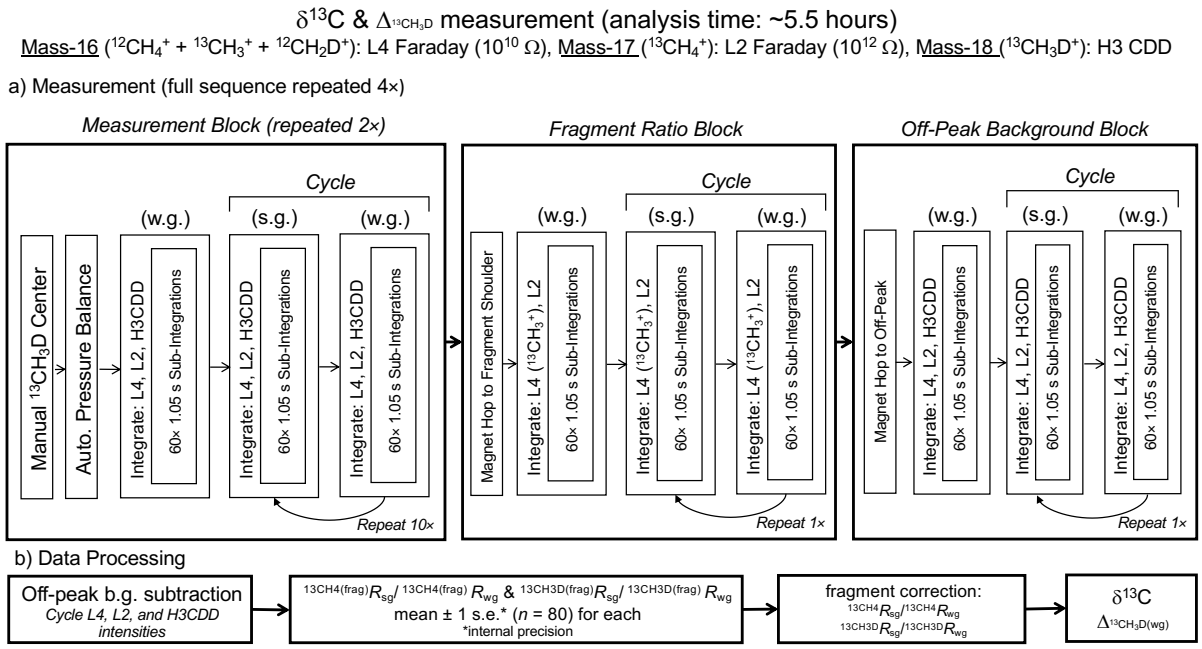

Figure SI.3(a-b): Combined $\delta^{13}C/\Delta_{13_{CH_3D}}$ flow-chart diagram illustrating measurement and data processing details. See main text (Section 2.4.2) for further details.

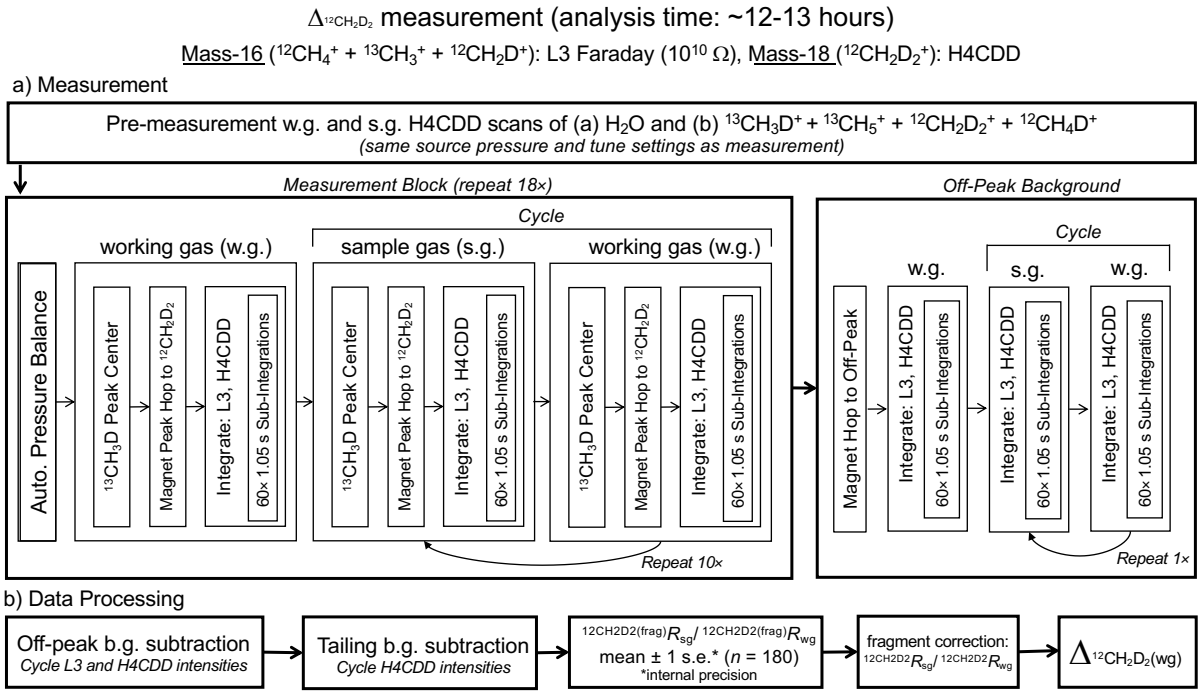

Figure SI.4(a-b): $\Delta_{12_{CH_2D_2}}$ flow-chart diagram illustrating measurement and data processing details. See main text (Section 2.4.3) for further details.

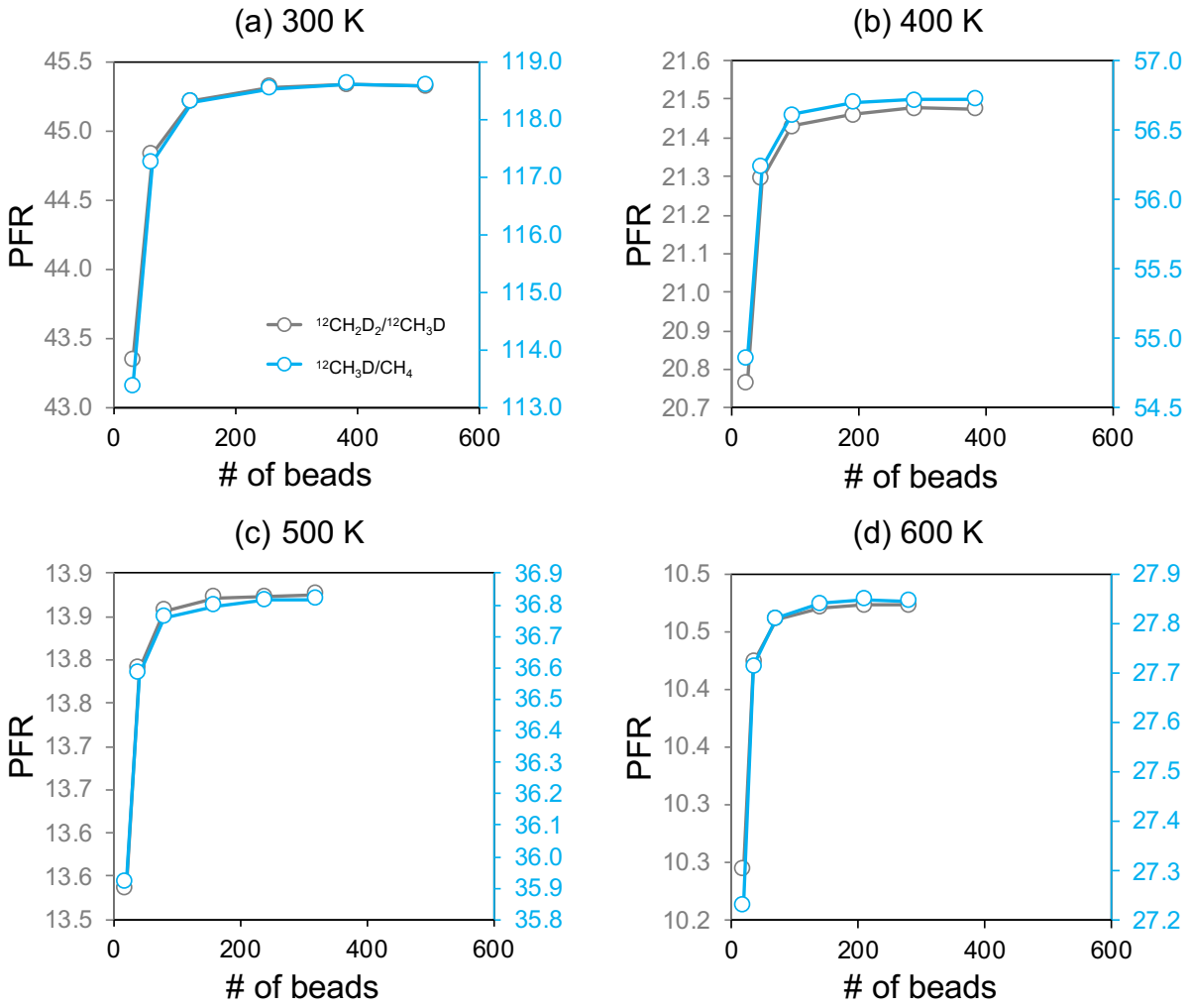

Figure SI.5(a-d): PIMC calculations: Convergence of $^{12}CH_3D/CH_4$ (blue) and $^{12}CH_2D_2/CH_3D$ (grey) PFR's with the number of beads. The number of beads for the other temperatures is determined by interpolation between the second last points on each panel. Data are from Table SI.1.

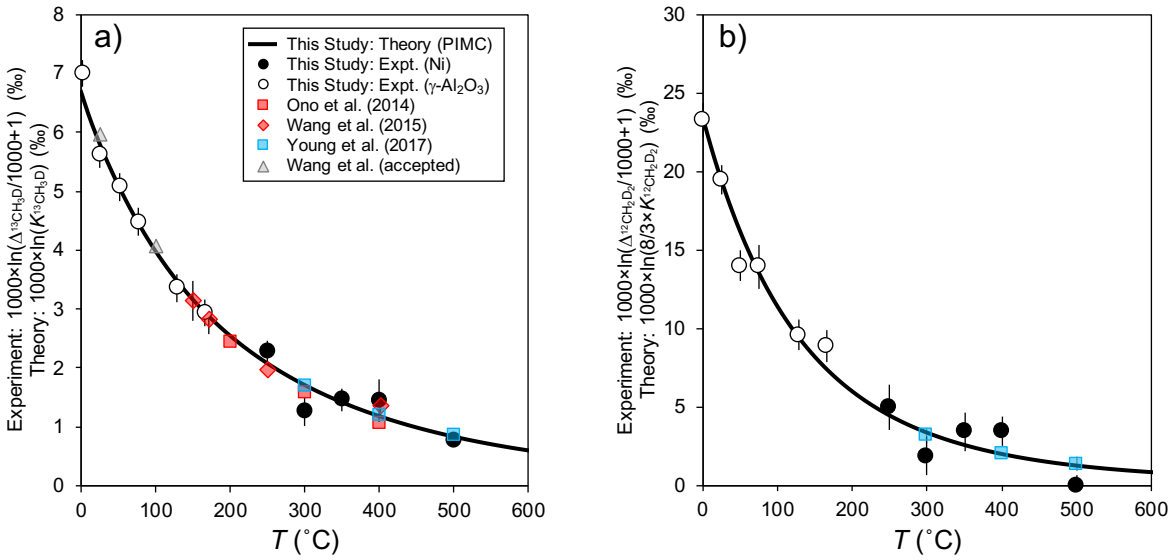

Figure SI.6(a-b): Comparison of the theoretical temperature dependence of $\Delta_{^{13}CH_3D}$ and $\Delta_{^{12}CH_2D_2}$ as given by $1000 \times \ln\left(K_{^{13}CH_3D}\right)$ and $1000 \times \ln\left(\frac{8}{3}K_{^{12}CH_2D_2}\right)$, respectively, based on the PIMC calculations of this study *vs.* $\Delta_{^{13}CH_3D}$ and $\Delta_{^{12}CH_2D_2}$ of samples equilibrated at known temperatures from this study and those of previous studies (Ono and co-workers, ref. 6; Wang and co-workers, ref. 16; Young and co-workers, ref. 7; Wang and co-workers, accepted, ref. 38; see reference list in main text). Data from Ono and co-workers (ref. 6) represents their "thermal equilibrium experiments" only and data from Wang and co-workers (ref. 16) were digitally extracted from their figure. Error bars for replicated experimental data points from this study represent either the ±1 s.e. of replicates (i.e., experimental reproducibility) or the expected ±1 s.e. based on the observed external precision of standards ($\sigma_{external}$) and the number of experimental replicates (i.e., $\sigma_{external}/\sqrt{n}$ where n = number of experimental replicates), whichever is larger. The error bar on one experimental data point from this study that has not been replicated (the $\Delta_{^{12}CH_2D_2}$ value at 350˚C) represents ±1 s.e. internal precision. Error bars from Ono and co-workers (ref. 6) and Young and co-workers (ref. 7) are ±1 s.e. and represent experimental reproducibility where replicated or internal precision of a single measurement when not replicated (error bars are generally smaller than symbols for these studies). Error bars for Wang and co-workers (ref. 16) represent 95% confidence intervals (see their paper for more detail). The error bars are not plotted for data points from Wang and co-workers (accepted, ref. 38) because the uncertainties on these determinations from their approach, to our knowedge, are not given. To first order, each laboratory appears to match the temperature dependence expected by theory (in this case, the PIMC calculations of the current study). The apparent strong agreement between laboratory groups is encouraging and likely indicates laboratory measurements are generally intercomparable despite different instrumentation, equilibration experiments, and theoretical calculations used for each laboratory's calibration.