Petersen Sierra V (Orcid ID: 0000-0002-4498-9776)
Saenger Casey (Orcid ID: 0000-0002-6345-1728)
Huntington Katharine W (Orcid ID: 0000-0003-0796-8850)
Kelson Julia Rae (Orcid ID: 0000-0002-0588-5018)
Bernasconi Stefano M. (Orcid ID: 0000-0001-7672-8856)
Kluge Tobias (Orcid ID: 0000-0001-7559-4737)
Olack Gerard (Orcid ID: 0000-0001-9196-8292)
Bajnai David (Orcid ID: 0000-0002-4053-5056)
Breitenbach Sebastian F.M. (Orcid ID: 0000-0001-9615-2065)
Fernandez Alvaro (Orcid ID: 0000-0003-4765-707X)
Hodell David A. (Orcid ID: 0000-0001-8537-1588)
Lohmann Kyger C (Orcid ID: 0000-0002-0709-9192)
Rosenheim Brad Erik (Orcid ID: 0000-0001-6141-4651)

Effects of Improved ^{17}O Correction on Inter-Laboratory Agreement in Clumped Isotope Calibrations, Estimates of Mineral-Specific Offsets, and Temperature Dependence of Acid Digestion Fractionation

S. V. Petersen[1], W. F. Defliese[2,3], C. Saenger[4], M. Daëron[5], K. W. Huntington[4], C. M. John[6], J. R. Kelson[4], S. M. Bernasconi[7], A. S. Coleman[8], T. Kluge[9], G. A. Olack[8], A. J. Schauer[4], D. Bajnai[10], M. Bonifacie[11], S. F. M. Breitenbach[12,13], J. Fiebig[10], A. B. Fernandez[7], G. A. Henkes[14], D. Hodell[13], A. Katz[11], S. Kele[15], K. C. Lohmann[1], B. H. Passey[1], M. Y. Peral[5], D. A. Petrizzo[16], B. E. Rosenheim[17], A. Tripati[3], R. Venturelli[17], E. D. Young[3], and I. Z. Winkelstern[18]

[1]University of Michigan, Ann Arbor, USA. [2]Texas A&M University, College Station, USA. [3]University of California - Los Angeles, Los Angeles, USA. [4]University of Washington, Seattle, USA. [5]Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France. [6]Imperial College London, London, United Kingdom. [7]ETH-Zurich, Zurich, Switzerland. [8]University of Chicago, Chicago, USA.[9]Heidelberg University, Heidelberg, Germany. [10]Goethe-Universität Frankfurt, Frankfurt am Main, Germany. [11]Institut de Physique du Globe de Paris, Paris, France. [12]Ruhr-Universität Bochum, Bochum, Germany. [13]Cambridge University, Cambridge, United Kingdom. [14]Stony Brook University, Stony Brook, USA. [15]Hungarian Academy of Sciences, Budapest, Hungary. [16]Jet Propulsion Laboratory, Pasadena, USA. [17]University of South Florida, St. Petersburg, USA. [18]Grand Valley State University, Allendale, USA.

Corresponding author: Sierra Petersen (sierravp@umich.edu)

Key Points:

- Updates to ^{17}O correction parameters and international standard compositions affect clumped isotopic compositions

- Reprocessing published calibration data using new parameters and consistent methodology slightly improves inter-laboratory agreement
- No evidence (within error) for mineral-specific offsets in calibration equation or temperature dependence of acid digestion fractionation

Abstract

The clumped isotopic composition of carbonate-derived CO_2 (denoted Δ_{47}) is a function of carbonate formation temperature, and in natural samples can act as a recorder of paleoclimate, burial or diagenetic conditions. The absolute abundance of heavy isotopes in the universal standards VPDB and VSMOW (defined by four parameters: R^{13}_{VPDB}, R^{17}_{VSMOW}, R^{18}_{VSMOW} and λ) impact calculated Δ_{47} values. Here, we investigate whether use of updated and more accurate values for these parameters can remove observed inter-laboratory differences in the measured T-Δ_{47} relationship. Using the updated parameters, we reprocess 14 published calibration datasets measured in 11 different laboratories, representing many mineralogies, bulk compositions, sample types, reaction temperatures, and sample preparation and analysis methods. Exploiting this large composite dataset (*n*=1253 sample replicates), we investigate the possibility for a "universal" clumped isotope calibration. We find that applying updated parameters improves the T-Δ_{47} relationship (reduces residuals) within most labs and improves overall agreement but does not eliminate all inter-laboratory differences. We reaffirm earlier findings that different mineralogies do not require different calibration equations, and that cleaning procedures, method of pressure baseline correction, and mass spectrometer type do not affect inter-laboratory agreement. We also present new estimates of the temperature dependence of the acid digestion fractionation for Δ_{47} ($\Delta^*_{25\text{-}X}$), based on combining reprocessed data from four studies, and new theoretical equilibrium values to be used in calculation of the empirical transfer function. Overall, we have ruled out a number of possible causes of inter-laboratory disagreement in the T-Δ_{47} relationship, but many more remain to be investigated.

Plain Language Summary

Measured stable and clumped isotope values are fundamentally tied to established compositions of international standard materials. When these standard compositions are updated, it impacts previously published isotope measurements such as those used to define the clumped isotope calibration relationship (the foundation for use of this isotopic proxy as a paleothermometer, recorder of burial history or past diagenetic conditions). Here we reprocess 14 published clumped isotope calibration studies using updated international standard compositions and identical data processing procedures to see if these changes would eliminate previously observed inter-laboratory discrepancies in clumped isotope calibration relationships. We find that this update tightens the clumped isotope calibration relationship within most laboratories and improves overall agreement between laboratories but does not eliminate all inter-laboratory differences. We also propose "best practices" for data processing and dissemination going forward. This study makes progress towards resolving discrepancies in clumped isotope calibration relationships between laboratories by eliminating a number of possible causes and moves the clumped isotope community closer towards our ultimate goal of applying this powerful new proxy routinely to exciting science questions.

Keywords: clumped isotopes, stable isotopes, calibration, standardization

[Index terms]
stable isotope geochemistry (1041), community standards (1904), data management, preservation, rescue (1912), data assimilation, integration and fusion (1910), general or miscellaneous geochemistry (1099)

1. Introduction

The carbonate clumped isotope paleothermometer was theorized (Schauble et al., 2006) and demonstrated (Ghosh et al., 2006) in the mid-2000's as a new method of measuring past temperatures using biogenic and inorganic carbonate materials. The biggest benefit of this new method is that it is based on a homogeneous equilibrium reaction (equilibrium distribution of isotopes within one material), and is not dependent on the isotopic composition of the fluid in which the carbonate formed like the traditional oxygen isotope paleothermometer (Ghosh et al., 2006). However, early calibration studies disagreed on the slope of the relationship between formation temperature and the clumped isotopic composition of synthetically precipitated carbonates (Dennis and Schrag, 2010; Ghosh et al., 2006). Following implementation of the 'absolute reference frame' (ARF), or 'carbon dioxide equilibrium scale' (CDES), correction method (Dennis et al., 2011), which adjusts clumped isotope measurements made in different laboratories or at different times onto a common scale using stochastic gases, gases equilibrated with water at known temperature and carbonates of established composition, the clumped isotope community found improved but not exact agreement between calibration studies. Just over a decade after the initial calibration of this novel proxy method, there is continued effort to resolve increasingly smaller inter-laboratory disagreements in calibration studies. For example, recent studies have shown that calibration slopes and intercepts may be biased by low numbers of sample points or replicates analyses, or by a limited temperature range investigated (Bonifacie et al, 2017; Fernandez et al., 2017; Katz et al., 2017; Kelson et al., 2017).

In order to calculate the clumped isotopic composition of CO_2 (Δ_{47}), the absolute abundance of heavy isotopes in the universal standards (VPDB, VSMOW) must be defined. This is done with four parameters: R^{13}_{VPDB} (the ratio of ^{13}C to ^{12}C in the VPDB carbonate standard), R^{17}_{VSMOW} and R^{18}_{VSMOW} (the ratio of ^{17}O or ^{18}O to ^{16}O in the VSMOW water standard), and λ (the slope of the triple oxygen line assumed to describe the fractionation of ^{17}O relative to ^{18}O in most natural carbonates). These four values are fundamental to the calculation of Δ_{47} from raw mass spectrometer outputs (raw voltages or currents) and are input very early in the calculation process. These values were explicitly defined to be 0.0112372, 0.0003799, 0.0020052, and 0.5164, respectively, in Huntington et al. (2009), as part of a step-by-step description of how to calculate Δ_{47} from raw voltages. This set of values has previously been called the "Gonfiantini" or "Santrock" parameter set after Gonfiantini et al. (1995) or Santrock et al. (1985), where they were previously defined together, although as individual parameter values, they are derived from separate earlier publications (Craig, 1957; Li et al., 1988; Baertschi, 1976; Matsuhisa et al. 1978, respectively).

With improved technology and understanding, our ability to absolutely define the abundance of heavy isotopes in these universal standard materials has improved. For example, a new value for λ has been defined based on the relationship seen in global meteoric waters, the reservoir from which many natural carbonates form (Luz & Barkan, 2010). These advancements were summarized by Brand et al. (2010), who put forward a new set of values (0.011180, 0.038475 (as calculated from $R^{18}{}_{VSMOW}$ using the new parameters), 0.0020052, 0.528, respectively) hereafter known as the "IUPAC" parameters (International Union of Pure and Applied Chemistry) and abbreviated as "Br" when needed. Recent studies have investigated the potentially large effects of changing these parameters on calculated Δ_{47} values, and several of them have advocated for use of the updated IUPAC parameter set over the prior Santrock/Gonfiantini (SG) parameter set (Daëron et al., 2016; Schauer et al., 2016). Due to the use of these fundamental parameters in calculating the Δ_{47} value of both gas and carbonate standards used in converting to the absolute reference frame, as well as for unknown carbonate samples, the predicted impact of updating to IUPAC parameters on a final sample Δ_{47} value is not straightforward and varies depending on the bulk composition ($\delta^{13}C$, $\delta^{18}O$) of the gas and carbonate standards relative to the bulk composition of the sample itself (Bernasconi et al., 2018; Daëron et al., 2016; Schauer et al., 2016).

Another source of uncertainty is the temperature dependence of the clumped isotope acid digestion fractionation. Δ_{47} values require correction to account for the effects of the removal of one oxygen atom during acid digestion (conversion of $CaCO_3$ to CO_2), which cause an increase in Δ_{47} values relative to corresponding Δ_{63} values in the carbonate phase (Bonifacie et al., 2017; Guo et al., 2009; Schauble et al., 2006). In practice, this is defined as the difference between Δ_{47} values determined from acid digestion at 25°C and a higher acid temperature (eg. 70°C, 75°C, 90°C, or 100°C). The temperature dependence of the acid digestion fractionation, denoted $\Delta^*_{25\text{-}X}$ (where X is the acid digestion temperature greater than 25°C) has been defined and redefined in a number of studies, in particular to determine whether mineralogical differences are required (Bonifacie et al., 2017; Defliese et al., 2015; Guo et al., 2009; Müller et al., 2017; Murray et al., 2016; van Dijk et al, 2019). The determination of $\Delta^*_{25\text{-}X}$ would also be affected by updating to IUPAC parameters.

An example of the effects of updating to IUPAC parameters on clumped isotope calibration data was demonstrated by Kelson et al. (2017). The authors synthetically precipitated inorganic carbonates at known temperatures using a variety of precipitation techniques, including those used by previous synthetic carbonate studies. The bulk composition of different samples in this study varied dramatically based on the precipitation method used (and therefore the source of carbon), which resulted in significant changes in Δ_{47} values and improved agreement between samples with the use of IUPAC parameters compared to Santrock/Gonfiantini parameters (removal of differences on the order of 0.05‰). This improved agreement was mainly the result of shifting intercept, as opposed to changes in slope, as predicted by Daëron et al. (2016).

Newly published clumped isotope studies have only begun adopting the new IUPAC parameters in the last year or two. Therefore, if one is limited to comparing published datasets, the scope of possible inter-laboratory comparisons is narrow, as demonstrated by two recent studies that compare data between 2-4 laboratories each (Bernasconi et al., 2018; Peral et al., 2018). In this study, we have gathered raw clumped isotope calibration data for which formation temperature is relatively well constrained (synthetic with error < ± 3°C, majority < ± 2°C, allowing higher error for biogenic and natural inorganic samples) from 14 publications, representing 11 different laboratories. We reprocessed each dataset with the same computer code, using both the old Santrock/Gonfiantini and the new IUPAC parameter sets. This compilation includes carbonates of variable mineralogy, bulk isotopic composition, and formation temperature, measured in 11 different laboratories using different acid reaction temperatures, sample cleaning procedures, and mass spectrometers (**Table 1**). We also reprocess data quantifying the temperature dependence of the acid digestion fractionation $\Delta^*_{25\text{-}X}$ from four studies and investigate the effects of using the IUPAC parameter set on this important value. With this comprehensive compilation, we seek to determine whether inter-lab discrepancies in the relationship between Δ_{47} and temperature may at least partially be accounted for by implementing the IUPAC parameter set and consistent data processing methods.

We find increased coherence in the Δ_{47} vs. temperature relationship within the majority of studies, taken individually, and somewhat improved inter-laboratory agreement when all studies are taken together and datasets are updated to the IUPAC parameter set and processed identically. Within the uncertainty of the data, we reaffirm a lack of evidence of mineralogy-dependent calibrations or temperature dependence of the acid digestion fractionation previously suggested by others (Bonifacie et al., 2017; Defliese et al., 2015). However, systematic discrepancies between laboratories and studies remain, and may be the result of choice of standardization scheme (gas vs. carbonate standards), kinetic effects during carbonate precipitation, differences in the preparation of orthophosphoric acid, design and operation of the preparation vacuum line, or in mass spectrometry methods. With this data reprocessing effort, we can rule out choice of parameter set and data treatment procedures as the cause of remaining inter-laboratory disagreement. This study also reinforces the growing consensus that the large discrepancies observed between the first two clumped isotope calibration studies are not present in more recent studies and remaining differences are much smaller. Continued improvement in standardization, replication, and calibration within the community means that more recent calibrations are more accurate, and for this reason we advise against the tendency that sometimes arises to default to the original Ghosh et al. (2006) calibration, despite its groundbreaking nature.

2. Data Selection and Reprocessing Methods

2.1 Data selection

We have gathered clumped isotope calibration data from 14 publications for use in this study, with a primary focus on calibration studies using synthetically precipitated carbonates that do not show evidence of kinetic effects, but also including for comparison a few “well-behaved” biogenic and natural inorganic carbonate datasets that approximate apparent equilibrium behavior and have well-constrained formation temperatures (**Table 1**). Certain biogenic carbonates (e.g. shallow water corals, nautiloids, brachiopods) and speleothems have been shown to deviate significantly from the generally established clumped isotope calibration line and thus were not included (Affek et al., 2008; Bajnai et al, 2018; Daëron et al., 2011; Davies and John, 2019; Dennis et al., 2013; Kluge and Affek, 2012; Saenger et al., 2012; Saenger et al., 2017; Spooner et al., 2016).

We restrict ourselves to studies that were carried out after the adoption of the absolute reference frame (Dennis et al., 2011), which was created to allow for the greatest inter-laboratory comparability. We allow conversion to the absolute reference frame using either gas standards (heated or equilibrated) driven to equilibrium through methods accepted by the community (Dennis et al., 2011), carbonate standards accessible to the whole community whose values have been adopted across multiple labs (so called "ETH standards"; Bernasconi et al., 2018; Breitenbach et al., 2018; Peral et al., 2018), or a combination of the two. We choose not to include older studies that require “retroactive” conversion to the absolute reference frame using in-house standards, although they could, in principle, be converted into the absolute reference frame using a secondary transfer function. "Established" values for in-house standards are unable to be verified by other groups, often unavailable processed with IUPAC parameters (reprocessing long-term data to re-compute established values is beyond the scope of this study), and introduces an additional source of uncertainty to our intercomparison that we eliminate with our data selection criteria.

In order to test whether universal adoption of IUPAC parameters causes convergence of data towards a single clumped isotope calibration equation, we seek out calibration studies using carbonates of varied mineralogy, including calcite, aragonite, dolomite, vaterite, siderite, and magnesite. We allow for different acid digestion temperatures (25°C, 70°C, 75°C, 90°C, and 100°C), reaction methodologies (common acid bath vs. individual Kiel vs. ‘McCrea-style’ reaction vessels (McCrea, 1950)), and cleaning procedures (helium flow through gas chromatograph (GC) column vs. custom-packed trap with helium carrier gas vs. static Porapak^{TM} (PPQ) trap with no carrier gas). We also include data collected on different mass spectrometer models (Thermo MAT253 vs. Isoprime 100), using different data acquisition techniques (traditional long integration time on single cup configuration vs. shorter integration time with peak hopping), and different pressure baseline (PBL) background correction methods (measured once per sample, once per acquisition, or no PBL correction at all). Details of sample preparation methodology for all included studies are summarized in **Table 1**.

These selection criteria result in the inclusion of 8 published inorganic carbonate calibration studies (Defliese et al., 2015; Fernandez et al., 2014; Garcia de Real et al., 2015; Kelson et al., 2017; Kluge et al., 2015; Passey and Henkes, 2012; Tang et al., 2014; Winkelstern et al., 2016). We also reprocessed 6 biogenic/natural inorganic calibration studies for comparison (Breitenbach et al., 2018; Henkes et al., 2013; Katz et al., 2017; Peral et al., 2018; Petrizzo et al., 2014; Wacker et al., 2014). In total, our analysis reprocessed 5448 unique sample and standard replicates measured in 11 different laboratories for a total of 263 individual samples (132 synthetic, 118 biogenic, 13 natural inorganic).

These criteria eliminate many calibration studies that partially or fully consist of data collected prior to the adoption of the absolute reference frame and/or do not follow the currently established reproducibility norms (sufficient number of standards analyzed, bracketing of unknown replicates by standard replicates, distribution of replicate analysis over many days). Eliminated studies are both synthetic (Dennis and Schrag, 2010; Falk and Kelemen, 2015; Ghosh et al., 2006; Stolper and Eiler, 2015; Zaarur et al., 2013) and biogenic/natural (Affek et al., 2008; Came et al., 2007, 2014; Daëron et al., 2011; Douglas et al., 2014; Eagle et al., 2010, 2013, 2015; Ghosh et al., 2007; Grauel et al., 2013; Kluge and Affek, 2012; Thiagarajan et al., 2011; Tripati et al., 2010; Zaarur et al., 2011) sample sets. Other studies were eliminated due to samples being affected by kinetic processes after a forced aragonite-to-calcite phase transition (Staudigel et al., 2016), insufficient constraint on formation temperature (low temperature magnesite from Garcia de Real et al., 2016; Came et al., 2017; Sample et al., 2017), lack of access to raw data (voltages/currents, or sample mean δ^{45}-δ^{49} values) needed for this exercise.

Our data selection criteria result in the elimination of the two original clumped isotope calibration studies (Dennis and Schrag, 2010; Ghosh et al., 2006), which show the greatest inter-study disagreement, as well as all other "steep slope" calibrations (Eagle et al., 2010; Ghosh et al., 2007; Tripati et al., 2010; Zaarur et al., 2013). Ghosh et al. (2006) in particular represents a very early study with few replicates per sample and insufficient standards measured, by today's standards and other "steep slope" calibrations all predate the adoption of the absolute reference frame. Although some of these early studies can be or have been converted into the absolute reference frame using in-house carbonate standards and secondary transfer functions (Daëron et al., 2016; Dennis et al., 2011; Zaarur et al., 2013), this is not ideal and inclusion of these studies would introduce additional, and difficult to constrain, uncertainty to our inter-lab comparison efforts, complicating our ability to test whether or not continued inter-laboratory disagreement can be explained by use of an outdated (Santrock/Gonfiantini) parameter set. However, based on the behavior of samples found here following parameter updates and the distribution of bulk compositions in samples from these two early studies, it is unlikely that the disagreement between the two early studies can be explained by use of Santrock/Gonfiantini parameters (as also concluded by Daëron et al., (2016)).

2.2 Reprocessing methods

Data processing workflow is summarized in **Figure 1.** Because the four fundamental parameters (R^{13}_{VPDB}, R^{17}_{VSMOW}, R^{18}_{VSMOW} and λ) are used in the earliest stages of the clumped isotope calculation, reprocessing must be carried out on very raw data. Where possible, we began with the raw mass spectrometer output (raw voltages or currents). When this was not possible, we began with the sample mean small delta values (δ^{45}, δ^{46}, δ^{47}, δ^{48}, and δ^{49}). Using data where raw voltages were available, we demonstrated that these two methods produce final $\delta^{13}C$, $\delta^{18}O$, and Δ_{47} values (in ‰) equivalent out to the 4th or 5th decimal place, well beyond the 3rd decimal place to which clumped isotope values are usually reported (**Text S1, Table S1**). If performed in the original study, background corrections ("pressure baseline correction" – Bernasconi et al., 2013; Fiebig et al., 2016; He et al., 2012; Peral et al., 2018; Petrizzo & Young, 2014; Rosenheim et al., 2013; Schauer et al., 2016; Venturelli & Rosenheim, 2018) are carried out prior to calculating the small delta values (**Table 1**).

Beginning with the small delta values (δ^{45} to δ^{49}), unless otherwise noted, all datasets are treated identically when calculating $\delta^{13}C$, $\delta^{18}O$, and raw Δ_{47} values. Data was processed using a modified version of the code in Huntington et al. (2009), rewritten in the statistical program 'R' [https://www.R-project.org/], and available in the **Supplementary Material**. $\delta^{18}O$ values of the analyzed CO_2 were calculated using the *optimize*() function in R and were not converted to carbonate $\delta^{18}O$ because it was not needed for this study. Some data contributors provided data analyzed using the freely available clumped isotope data analysis software "Easotope" (John and Bowen, 2016), already processed using both parameter sets (achievable in this program in only a few clicks). Easotope produces final values nearly identical to those produced by the R code (**Text S1**), but in order to eliminate as many variables in data processing as possible, Easotope data was reprocessed from sample mean δ^{45} to δ^{49} values as well. This resulted in Δ_{47} values within 0.001‰ of Easotope output and highlighted the effects of rounding intermediate values. In the R code, $\delta^{13}C$, $\delta^{18}O$, and raw Δ_{47} values were all calculated using a single parameter set (either Santrock/Gonfiantini (SG) or IUPAC (Br)), unlike the original Huntington et al. (2009) code which used $\delta^{13}C$ and $\delta^{18}O$ values exported by the Thermo data analysis program ISODAT instead of calculating them explicitly. ISODAT uses a set of parameters differ slightly from the Santrock/Gonfiantini parameter set described above, with $R^{17}_{VSMOW} = 0.0004023261$ (Santrock et al., 1985) instead of 0.0003799 (Huntington et al., 2009; Li et al., 1988).

To transition raw Δ_{47} values into the absolute reference frame, we reproduce the correction methods of the original study as closely as possible. We select the same reference frame correction intervals or 'windows', and use either a fixed interval, moving window, or combination approach, whichever was originally used. Given the highly variable behavior and stability of different mass spectrometers, we assume the authors of the original study knew best and had optimized the correction methods for their individual machine and study interval. Study-specific notes on correction methods can be found in the **Text S2** and are summarized in **Table 1**.

Transition of raw Δ_{47} values into the absolute reference frame (ARF) (Dennis et al., 2011) requires calculation of three parameters: SlopeEGL (slope of equilibrium gas line(s) in δ^{47} vs. Δ_{47} space – either heated gases, CO_2-H_2O equilibrations, or a combination of the two fit together), SlopeETF and IntETF (slope and intercept of the empirical transfer function). The empirical transfer function (ETF) plots the intercepts of the equilibrium gas lines (or carbonate standards corrected for SlopeEGL) against the "true" Δ_{47} value for the corresponding equilibration/formation temperature (1000°C for heated gases, typically 0-60°C for CO_2-H_2O equilibrations, defined value for carbonate standards previously established relative to gas standards). This "true" Δ_{47} value or theoretical equilibrium Δ_{47} value (hereafter $\Delta_{47\text{-TE}}$) was originally defined as the thermodynamic equilibrium value at the corresponding equilibration temperature, and was approximated using equation A2 of Dennis et al. (2011), a 4th order polynomial fit through thermodynamic calculations of Wang et al. (2004). Due to rounding of universal constants used in the original thermodynamic calculations (Planck's constant, speed of light, and Boltzmann constant with only 0-2 decimal places in scientific notation instead of 8-9 as they are best defined (Fisher & Ullrich, 2016)) and in conversion between Celsius and Kelvin (using +273 instead of +273.15), as well as inherent uncertainty in using a polynomial fit approximation, the $\Delta_{47\text{-TE}}$ values assumed in different studies differ slightly (**Text S3, Figure S1**). To eliminate this variable in our inter-study comparison, we recalculated the thermodynamic equilibrium values using long-format versions of the universal constants, the correct conversion between Celsius and Kelvin, and a higher-order (7th-order) polynomial fit with long-format decimal coefficients. We reprocess all calibration data using both the older values ("WD" for Wang/Dennis) and new values ("P" for Petersen et al., this study) (**Figure 1**). We provide a read-off table for the new $\Delta_{47\text{-TE}}$ values at many temperatures between -12°C and 1000°C, as well as the 7th order polynomial fit through this new data to be used for temperatures not included in the read-off table (**Table S2**). The choice of $\Delta_{47\text{-TE}}$ values changes the SlopeETF and IntETF ARF parameters but not SlopeEGL.

In the case of studies using carbonates to convert data into the absolute reference frame, we use published $\Delta_{47\text{-TE}}$ values where possible and otherwise (e.g. for in-house standards) rely on author-provided values, established within the absolute reference frame using gases within an individual lab. Of labs using carbonate standards exclusively (ETH-Zurich, LSCE, Cambridge) or in-combination with gas standards (Imperial) to create a reference frame, most use the "ETH standards", four pure carbonates with varying bulk and clumped isotopic compositions created and provided to the community by the ETH-Zurich lab (Bernasconi et al., 2018; Meckler et al., 2014; Müller et al., 2017). For Santrock/Gonfiantini reprocessing, we use $\Delta_{47\text{-TE}}$ values published by Müller et al. (2017) for the ETH carbonate standards (ETH-1 = 0.265‰, ETH-2 = 0.267‰, ETH-3 = 0.703‰, ETH-4 = 0.522‰). For IUPAC parameter reprocessing, we use $\Delta_{47\text{-TE}}$ values from Bernasconi et al. (2018) (ETH-1 = 0.258‰, ETH-2 = 0.256‰, ETH-3 = 0.691‰, ETH-4 = 0.507‰). Use of identical values across studies using carbonates to convert to the absolute reference frame has been shown to improve inter-laboratory agreement (Bernasconi et al., 2018). A benefit of using exclusively carbonate standards is that even if $\Delta_{47\text{-TE}}$ values used initially are incorrect, using the same values to correct data from all laboratories allows for direct comparison within a single framework. This should also be the case when using exclusively gas standards and was the initial definition of the absolute reference frame (Dennis et al., 2011). However, gas standards do not account for the effects of acid digestion on sample unknowns. Issues could also arise if incorrect $\Delta_{47\text{-TE}}$ values of carbonate standards are combined with defined $\Delta_{47\text{-TE}}$ values of gas standards. In this case, our choice to use identical $\Delta_{47\text{-TE}}$ values for all laboratories using the ETH standards (exclusively or in combination with gas standards) is meant to eliminate another potential variable.

The last step towards calculating a final Δ_{47} value is correction for the reaction-temperature-dependent fractionation induced by loss of one oxygen in conversion from calcium carbonate ($CaCO_3$) to CO_2. The full acid fractionation factor ($\Delta^*_{\text{mineralogyX}}$ for a given mineralogy and acid temperature X; as defined by Bonifacie et al., 2017) relates the measured Δ_{47} value of extracted CO_2 and the Δ_{63} value in the original solid carbonate (Schauble et al., 2006). In practice, instead of using $\Delta^*_{\text{mineralogyX}}$ to correct CO_2 Δ_{47} values to a carbonate Δ_{63} value, measured Δ_{47} values are corrected relative to a reference reaction at 25°C by applying a fixed value, colloquially and ambiguously known as the "acid fractionation factor", that is a function of acid temperature X ($\Delta^*_{25\text{-X}}$; as defined by Bonifacie et al., 2017). Without applying any acid fractionation correction, Δ_{47} values for samples reacted at the same temperature can be directly compared, but a choice of $\Delta^*_{25\text{-X}}$ (or $\Delta^*_{T1\text{-}T2}$ between any two reaction temperatures) is necessary to compare samples reacted at different acid temperatures. A variety of theoretical (Guo et al., 2009) and experimental (Bonifacie et al., 2017; Defliese et al., 2015; Guo et al., 2009; Henkes et al., 2013; Kelson et al., 2017; Müller et al., 2017; Murray et al. 2016; Passey et al., 2010; Tripati et al., 2015; Wacker et al, 2013; Winkelstern et al., 2016; van Dijk et al., 2019) values for $\Delta^*_{\text{11ineralogy}}$ and $\Delta^*_{25\text{-X}}$ have been put forward. Many of these experimental studies have found similar behavior across differing mineralogies (Bonifacie et al., 2017; Defliese et al., 2015; Kelson et al., 2017; Wacker et al, 2013; Winkelstern et al., 2016), with either non-distinguishable values ($\Delta^*_{\text{dolomite90}}$ values within error of theoretical $\Delta^*_{\text{calcite90}}$ values (Bonifacie et al., 2017)), overlapping calibration data (Bonifacie et al., 2017; Defliese et al., 2015; Kelson et al., 2017; Wacker et al, 2013; Winkelstern et al., 2016), or identical $\Delta^*_{25\text{-X}}$ values across multiple mineralogies (Defliese et al., 2015). However, a few studies have found mineralogy-specific behavior in Δ_{47} (Müller et al., 2017; Murray et al. 2016; Tripati et al., 2015; van Dijk et al., 2019), such as Müller et al. (2017) who found a measurable difference on the order of 0.05‰ between calcite, aragonite and dolomite ($\Delta^*_{\text{calcite70}}$ vs. $\Delta^*_{\text{aragonite70}}$ vs. $\Delta^*_{\text{dolomite70}}$). **Figure 1** includes a summary of abbreviations and subscripts relating to, among other things, acid digestion fractionation corrections.

In order to eliminate another variable from our inter-study comparison (and in accordance with our findings here, discussed below), we choose to use the same $\Delta^*_{25\text{-X}}$ values for all mineralogies, although this assumption is explored further in this study. We derive our $\Delta^*_{25\text{-X}}$ values by reprocessing and combining data from four studies (Defliese et al., 2015; Henkes et al., 2013; Kelson et al., 2017; Kluge et al., 2015). Because our primary objective was reprocessing calibration datasets, we limited ourselves to $\Delta^*_{25\text{-X}}$ data contained in the 14 studies listed in Table 1. Unfortunately, none of the studies showing mineralogical differences in Δ_{47} or $\Delta^*_{25\text{-X}}$ (Müller et al., 2017; Murray et al. 2016; Tripati et al., 2015) were included. New $\Delta^*_{25\text{-X}}$ values can be found in **Table 2** and **Table S3**. In practice, we apply four slightly different sets of $\Delta^*_{25\text{-X}}$

values in this intercomparison, depending on the appropriate combination of SG vs. IUPAC parameters and W/D vs. P $\Delta_{47\text{-TE}}$ values used to calculate sample $\Delta_{47\text{-RF}}$. In some instances, we will compare $\Delta_{47\text{-RF}}$ values without adding $\Delta^*_{25\text{-X}}$ to isolate the effects of changing parameter sets or $\Delta_{47\text{-TE}}$ values. Final calibration data can easily be presented relative to a reference reaction at 90°C by subtracting the $\Delta^*_{25\text{-}90}$ value from final values (although this implicitly includes some acid fractionation correction for samples not reacted at 90°C).

After all these steps, the result is a Δ_{47} value converted into the absolute reference frame and corrected for the acid digestion fractionation ($\Delta_{47\text{-RFAC}}$, reference frame acid corrected), processed using either Santrock/Gonfiantini parameters (SG) or IUPAC parameters (Br), with use of either Wang/Dennis $\Delta_{47\text{-TE}}$ values (WD) or Petersen et al. values (P) from this study (**Figure 1**). Datasets will be referred to by the first author's last name as opposed to a full citation in figures and discussion where new conclusions are drawn based on reprocessing. A full citation will be used where we intend to reference an insight or conclusion from the original study.

2.3 Treatment of uncertainty

The measurement of the doubly-substituted, mass-47 isotopologue of CO_2 central to the clumped isotope method ($^{13}C^{18}O^{16}O$) is conducted near the shot-noise limits of modern mass spectrometry technology. To combat this, users generally measure each sample unknown a minimum of 3 times (*n*=3 replicates of ~3-8mg $CaCO_3$ each) or more (*n*>7 replicates of ~0.1 mg $CaCO_3$ with a Kiel device (Schmid & Bernasconi, 2010) or *n*=4-6 replicates of ~1-1.5 mg $CaCO_3$ (Petersen and Schrag, 2014)). The final Δ_{47} value is then taken as the mean of these *n* replicates and the error is the standard error on the mean (1SE = 1sd/SQRT(*n*), 'internal error'). Depending on when these replicates were run relative to each other, this internal error may underestimate the true variability of replicates of the same sample over time (long-term, 'external error') (Fernandez et al., 2017). External error is better captured by the long-term reproducibility of carbonate standards, run many times over the measurement intervals. External error for a sample unknown measured *n* times is then taken as the standard error calculated using the average standard deviation of the carbonate standards measured many times (1extSE = 1sd_carbstds/SQRT(*n*)). As was recently recommended by others (Bonifacie et al, 2017; Fernandez et al., 2017; Kelson et al., 2017), we present both internal and external 1SE values on sample means, but choose to use the external error as representing the overall uncertainty. In particular, we preferred the external error in this case because it is unchanged across the transition from Santrock/Gonfiantini to IUPAC parameters, leading to identical weighting of samples in comparable regressions, unlike in Levitt et al. (2018).

2.4 Linear Regression Methods and Inter-Laboratory Comparison of Δ_{47}-T Calibrations

The temperature sensitivity of reprocessed data was evaluated via a Monte Carlo least squares regression approach to take into account error in estimated formation temperature and measured Δ_{47}. Specifically, a Δ_{47} and formation temperature pair was selected from within the

1SE (external) uncertainty bounds of each individual replicate assuming normal distributions, and a linear model was fit to these data. This approach has the advantage of giving greater weight to samples with greater replication. The routine was repeated 10,000 times to build a distribution of calibration slopes and intercepts that accounts for uncertainty in Δ_{47} measurements and formation temperature estimates. Inputs to this analysis are 1) assumed Gaussian error on each Δ_{47} replicate, based on 60-120 cycles of the mass spectrometer and 2) assumed Gaussian errors in formation temperature as reported by initial authors (independent of measured Δ_{47}). Therefore, the output distributions in slope and intercept are also expected to be Gaussian. The mean and standard error of the 10,000 iterations was used as the best estimate of each parameter of the linear model. This approach was applied to generate Δ_{47}-temperature relationships for all combinations of parameter sets (SG and IUPAC), $\Delta_{47\text{-TE}}$ values (WD and P), and $\Delta^*_{25\text{-X}}$ relationships (Defliese et al., 2015, and our updated compilation, using the appropriate parameters and $\Delta_{47\text{-TE}}$ values). Regressions were performed on subsets of data that included only synthetic data, and data with formation temperatures below 100ºC (**Table 3**, **Table S4**).

For each individual pair of calibration studies, we assess whether the observations are likely to reflect a single underlying T-Δ_{47} relationship by performing an analysis of covariance (ANCOVA), modified to account for the assigned uncertainties in temperature and Δ_{47} observations. In a first step, we compute two independent weighted orthogonal distance regression lines, one for each of the two studies. This yields best-fit values and model standard errors for both slopes and both intercepts. Note that in order to better account for potential unrecognized sources of error, for studies with a chi-squared value larger than the degree of freedom in the model, slope and intercept standard errors are conservatively scaled by the square root of the reduced chi-square, $\sqrt{\chi^2/(N-2)}$, N being the number of observations. We may then estimate the probability (P_s) for the null hypothesis that the two slopes are identical, taking into account the model standard errors. If the slopes are found to be statistically indistinguishable (at a 95% confidence level), the observations from both datasets are jointly fit to a new regression model with two parallel lines. If the difference in the intercept values of these two lines is statistically indistinguishable from zero (with a null-*p* value of P_i), the hypothesis that the two datasets reflect a single underlying relationship between Δ_{47} and T cannot be excluded at that confidence level.

We assess community convergence in two more ways. We compare single-study Deming regressions (taking into account error in x and y, or formation temperature and Δ_{47} in this case) and calculate the range in Δ_{47} predicted for a formation temperature of 25°C and, similarly, the range in temperature predicted for a Δ_{47} value of 0.700‰.

3. Results and Discussion

3.1 Effects of updating to new theoretical equilibrium Δ_{47} ($\Delta_{47\text{-TE}}$) values on sample $\Delta_{47\text{-RF}}$

The second step in conversion of Δ_{47} values into the absolute reference frame requires defining the 'true' Δ_{47} values for gas and carbonate standards. These are defined as the thermodynamic equilibrium Δ_{47} value of CO_2 at the corresponding temperature ($\Delta_{47\text{-TE}}$) and were defined by Dennis et al. (2011) based on thermodynamic calculations of Wang et al. (2004). Due to the order of the polynomial fit and the imprecise conversion from Celsius to Kelvin implemented by Dennis et al. (2011) in initially defining these values, $\Delta_{47\text{-TE}}$ values could vary, even for the same equilibrium temperature, depending on whether they were taken from the table of raw theoretical values (Dennis et al., 2011, Data), the published summary table (Dennis et al., 2011, Table 1), or were calculated anew using the polynomial fit (Dennis et al., 2011, Equation A2). This inconsistency in the definition of $\Delta_{47\text{-TE}}$ values used by different laboratories contributes to the observed scatter in data between studies.

As part of this study, we recalculated the thermodynamic equilibrium values following Wang et al. (2004), using long-form versions of universal constants, precise conversion from Celsius to Kelvin, and updated IUPAC parameters (**Text S3**). New $\Delta_{47\text{-TE}}$ values (Petersen, this study, abbreviated as P) are lower than published Wang/Dennis (WD) values across the entire temperature range spanning 0 to 1000°C by 0 to 0.006‰, with the largest discrepancies occurring around 25°C, near the temperature at which many labs produce equilibrated gases (**Figure S1**). Use of IUPAC vs. Santrock/Gonfiantini parameters had a very minor effect on calculated $\Delta_{47\text{-TE}}$ values (0-0.0005‰), with the dominant contributor to the change being use of long-form versions of universal constants (**Text S3**). Updating to new $\Delta_{47\text{-TE}}$ values causes a decrease in final sample $\Delta_{47\text{-RFAC}}$ values ($\Delta\Delta_{47\text{-RF}}$ WD-P) of 0 to 0.006‰, with similar behavior across all reprocessed datasets (**Figure 2**). Excluding data that show no difference in $\Delta_{47\text{-RFAC}}$ when new Petersen $\Delta_{47\text{-TE}}$ values are used (data from Breitenbach et al. (2018), Peral et al. (2018) and a small number samples from the Kluge et al. (2015) and Garcia de Real et al. (2016)), the mean shift in $\Delta_{47\text{-RF}}$ is 0.003 ± 3E-5‰ (1SE). These samples that have $\Delta\Delta_{47\text{-RF}}$ (WD-P) of zero (difference in $\Delta_{47\text{-RF}}$ calculated using WD vs. P $\Delta_{47\text{-TE}}$ values; **Figure 1**) are all from measurement sessions that used exclusively 1000°C heated gases and carbonate standards (no equilibrated gases) or carbonate standards alone. $\Delta_{47\text{-TE}}$ values for carbonate standards were not changed between WD and new ETF calculations, and the $\Delta_{47\text{-TE}}$ value for heated gases (1000°C) is identical to the 4th decimal place in both WD and new calculations. In reality, $\Delta_{47\text{-TE}}$ values for all carbonate standards should be updated as well because their established $\Delta_{47\text{-TE}}$ values are inherently tied to gas standards using WD $\Delta_{47\text{-TE}}$ values via the absolute reference frame, but reprocessing at this deeper level is beyond the scope of this study. Bernasconi et al. (2018) argues that as long as a consistent set of $\Delta_{47\text{-TE}}$ values are used by all laboratories, the benefits of the carbonate-based standardization will be realized, even if $\Delta_{47\text{-TE}}$ values are not correct.

The effect of using slightly different $\Delta_{47\text{-TE}}$ values in original data corrections due to varied readings of Dennis et al. (2011) (i.e. values taken from table vs. calculated from polynomial equation) is likely comparable to the order of magnitude of $\Delta\Delta_{47\text{-RF}}$ (WD-P) (0-0.006‰). Although we did not strictly quantify the effects on $\Delta_{47\text{-RF}}$ of establishing a consistent set of $\Delta_{47\text{-TE}}$ values (WD or P) compared to using variable $\Delta_{47\text{-TE}}$ values between studies, we expect the order of magnitude to be the same (~0-0.006‰), much less than the effects of changing from Santrock/Gonfiantini to IUPAC parameters (–0.025‰ to +0.044‰, see below). Although small, this magnitude of variability is relevant compared to the size of remaining differences between clumped isotope data from different laboratories and studies (see below).

Beyond issues created by varied readings of Dennis et al. (2011), additional definitions/calculations of $\Delta_{47\text{-TE}}$ values exist (Cao and Liu, 2012; Webb and Miller, 2013). Hill et al. (2014) calculated $\Delta_{47\text{-TE}}$ values using a more accurate and comprehensive method than used here, excluding some of the approximations and assumptions present in both Wang et al. (2004) and our $\Delta_{47\text{-TE}}$ values. However, all these authors only explicitly report $\Delta_{47\text{-TE}}$ values at a few temperatures, making the results less easily accessible, and as a consequence, none of these $\Delta_{47\text{-TE}}$ values have been widely adopted. In an effort towards establishing community best practices and improved inter-laboratory agreement, we recommend using the single set of $\Delta_{47\text{-TE}}$ values calculated as part of this study (listed in an easily accessible read-off table in **Table S2**) instead of the polynomial fit equation in Dennis et al. (2011), which will bring a higher level of consistency to past and future studies.

3.2 Effects of updating from Santrock/Gonfiantini to IUPAC parameters on sample $\Delta_{47\text{-RF}}$

Across all fourteen reprocessed datasets, using IUPAC parameters instead of Santrock/Gonfiantini parameters causes a change in reference-frame-acid-corrected Δ_{47} values ($\Delta\Delta_{47\text{-RF}}$ SG-Br) of –0.025‰ to +0.044‰ (**Figure 3a**) with changes varying in magnitude by laboratory and study. In general, $\Delta\Delta_{47\text{-RF}}$ (SG-Br) values show a positive correlation with sample δ^{13}C (**Figure 3b**) and a weaker negative correlation with sample δ^{18}O (**Figure 3c**), although the strength of the correlation varies by laboratory and study. Multiple studies (Defliese, Kelson, Henkes/Passey, Wacker, and Katz) show significant positive correlations ($r^2 > 0.75$, p-value < 0.05) between $\Delta\Delta_{47\text{-RF}}$ (SG-Br) values and sample δ^{13}C (**Figure S2**). A significant negative correlation ($r^2 > 0.75$, p-value < 0.05) with sample δ^{18}O is seen in four studies (Winkelstern, Garcia, Fernandez/Tang, and Peral) (**Figure S3**). The largest $\Delta\Delta_{47\text{-RF}}$ (SG-Br) values (up to +0.044‰) are seen in the studies from University of Michigan (Defliese/Winkelstern) (**Figure 3a**).

The magnitude of $\Delta\Delta_{47\text{-RF}}$ (SG-Br) is difficult to predict without fully reprocessing data because it depends on both the bulk composition (δ^{13}C and δ^{18}O) of the samples and gas/carbonate standards used to calculate the ARF parameters. In particular, the larger the difference between sample and standard compositions, the larger $\Delta\Delta_{47\text{-RF}}$ (SG-Br). (Note: This is different than Equation 10 in Daëron et al. (2016) which can be used to accurately predict changes in raw Δ_{47}). **Figure S4** summarizes the bulk compositions of samples and gas/carbonate standards used in each study. Bernasconi et al. (2018) showed that the magnitude of $\Delta\Delta_{47\text{-RF}}$ (SG-Br) is larger in laboratories that use gas standards with nearly constant δ^{13}C and variable δ^{18}O compared with laboratories that use gas standards of variable δ^{13}C and δ^{18}O owing to the fact that the effect of changing parameters is of opposite sign relating to δ^{13}C and δ^{18}O, confirming a similar finding from a synthetic data study (Daëron et al., 2016, Figure 6). Daëron et al. (2016)

also found that where sample $\delta^{13}C$ is equal to gas standard $\delta^{13}C$, $\Delta\Delta_{47\text{-RF}}$ (SG-Br) will be zero. This is demonstrated most clearly in datasets from Henkes/Passey and Kelson (**Figure 3b, Figure S2**). In these studies, heated and equilibrated gases were all created from a single tank imparting a single $\delta^{13}C$ composition on all gas standards (tank $\delta^{13}C \approx -5.5$‰ for Henkes/Passey and ≈ -10‰ for two of three measurement sessions in Kelson (the third measurement session also included standards with a $\delta^{13}C$ composition of –35.5‰)) (**Figure S4**). The composition of the tank corresponds to the points of zero $\Delta\Delta_{47\text{-RF}}$ (SG-Br) for each study (**Figure 3b, Figure S2**). Katz et al. (2017) did not use a single tank, but the $\delta^{13}C$ compositions of gas standards varied over a very narrow range around 0‰ (**Figure S4**), so a positive correlation with sample $\delta^{13}C$ composition intersecting zero $\Delta\Delta_{47\text{-RF}}$ (SG-Br) around 0‰ is still visible.

3.3 Effects of updating from Santrock/Gonfiantini to IUPAC parameters on the temperature dependence of the acid digestion fractionation ($\Delta^*_{25\text{-X}}$)

Changes in $\Delta^*_{25\text{-}90}$ are small when updating to IUPAC parameters. This is not surprising, because quantifying $\Delta^*_{25\text{-X}}$ is done by measuring the same sample at multiple reaction temperatures and, although the reaction temperature has an impact on both Δ_{47} and $\delta^{18}O$, the change in bulk composition is minor between samples reacted at 90°C and 25°C, meaning that changes in Δ_{47} due to updating the parameter set will be very similar for all replicates of that sample, assuming they were converted into the absolute reference frame using similar gas standards and/or carbonates.

To date, only three studies have published data relevant to the acid digestion fractionation calculated using IUPAC parameters (Kelson et al., 2017; Müller et al., 2017; van Dijk et al., 2019). Three other studies reprocessed here in addition to Kelson et al. (2017) include samples reacted at multiple acid temperatures, either explicitly to measure the temperature dependence of the acid digestion fractionation ($\Delta^*_{25\text{-X}}$) (Defliese et al., 2015; Henkes et al., 2013) or due to changing lab practices (Kluge et al., 2015). Relevant reprocessed data from all four studies were combined to assess the effects of updating from SG to IUPAC parameters on calculated $\Delta^*_{25\text{-X}}$ values. Unfortunately, raw data from some of the earliest analyses from Defliese et al. (2015) have been lost, including many of the 25°C replicates. Nevertheless, taken together, these four studies include 151 replicates covering the three most common mineralogies (calcite (*n*=87), aragonite (*n*=48), dolomite (*n*=16)). This includes good representation at both 25°C (*n*=52) and 90°C (*n*=67), allowing for robust assessment of $\Delta^*_{25\text{-}90}$. A single vaterite sample from Kluge et al. (2015) having only two replicates each at 70°C and 90°C was excluded from analysis due to lack of sufficient replicates and acid temperature range for that mineralogy. This represents the largest dataset to date directed at resolving the temperature sensitivity of the clumped isotope acid digestion fractionation.

Looking at $\Delta_{47\text{-RF}}$ (Br,P), we find values of +0.066‰, +0.072‰, +0.088‰, and +0.098‰ for reactions at 70°C, 75°C, 90°C, and 100°C, respectively ($\Delta^*_{25\text{-}70}$, $\Delta^*_{25\text{-}75}$, $\Delta^*_{25\text{-}90}$, $\Delta^*_{25\text{-}100}$). This corresponds to a temperature sensitivity of the acid digestion fractionation of -0.0010/°C, identical to the theoretical prediction of Guo et al. (2009). We find that within this dataset, updating from SG to IUPAC parameters causes a reduction in $\Delta^*_{25\text{-}90}$ of 0.006‰ ($\Delta^*_{25\text{-}90}$ = +0.095‰ for SG,WD vs. +0.089‰ for Br,WD, all mineralogies combined, *n*=151) and switching from Wang/Dennis to Petersen $\Delta_{47\text{-TE}}$ values causes an additional reduction of 0.001‰ ($\Delta^*_{25\text{-}90}$ = +0.088‰ for Br,P) (**Table 2**). The error on these $\Delta^*_{25\text{-}90}$ values (±0.006‰ for all mineralogies combined) is such that the overall change of 0.007‰ is barely statistically significant. Similar reductions in the $\Delta^*_{25\text{-}90}$ of 0.005-0.007‰ are seen for calcite and aragonite, when treated separately, suggesting the direction of change is robust, but note that the error on these subsets of data is larger (±0.008‰ for calcite, ±0.011‰ for aragonite). In contrast, the dolomite data, which is made up 16 analyses of a single sample from a single lab, shows a total reduction in $\Delta^*_{25\text{-}90}$ of only 0.001±0.010‰ from SG,WD to BR,P.

We find that regardless of parameter set used, calcite, aragonite, and dolomite produce $\Delta^*_{25\text{-X}}$ values that are equivalent within error (**Table 2**) supporting previous findings by some that multiple mineralogies can use the same $\Delta^*_{25\text{-X}}$ values (Bonifacie et al., 2017; Defliese et al., 2015; Kelson et al., 2017; Kluge and John, 2015; Wacker et al., 2013; Winkelstern et al. 2016). We must note, however, that none of the data that initially suggested large mineralogical differences in $\Delta^*_{25\text{-X}}$ was included in this reprocessing effort. This dataset is not able to interrogate $\Delta^*_{\text{mineralogyX}}$ values (difference between Δ_{47} of CO_2 and Δ_{63} of solid carbonate), which some have suggested shows mineralogical differences as well (Müller et al., 2017; van Dijk et al., 2019).

For internal consistency and to more completely quantify the effects of parameter change, we apply the same set of $\Delta^*_{25\text{-X}}$ values for all samples, regardless of mineralogy. When choosing which $\Delta^*_{25\text{-X}}$ values to use, we recommend following a similar practice and selecting the $\Delta^*_{25\text{-X}}$ value corresponding to not only your acid reaction temperature X, but also the appropriate ^{17}O correction parameters and $\Delta_{47\text{-TE}}$ values. Calculated $\Delta^*_{25\text{-X}}$ values for different parameter sets and mineralogies can be found in **Table 2** and **Table S3**.

3.4 Composite synthetic calibration using Santrock vs. IUPAC parameters

Considering only synthetic carbonates (studies 1-8 in **Table 1,** with natural dolomites from Winkelstern et al. (2016) and all magnesites from Garcia de Real et al. (2016) excluded), disagreement between laboratories is reduced after updating to IUPAC parameters but is not eliminated (**Figure 4**). Visually, this improved agreement is most apparent in the high temperature range, where the Winkelstern, Passey, and Kluge data align better when IUPAC parameters are used to process data (**Figure 4c, 4d**). Statistically, this is evidenced by reduced total residuals listed in **Figure 4c and 4d** and slightly improved r^2 values listed in **Table 3**.

Using synthetic carbonate samples only (*n*=451 replicates), processed using IUPAC parameters, new $\Delta_{47\text{-TE}}$ values, and new $\Delta^*_{25\text{-X}}$ values to scale everything to a 25°C acid temperature, the regression equation produced by Monte Carlo sampling is:

$$\Delta_{47-RFAC(Br,P,newAFF)} = (\mathbf{0.0383} \pm \mathbf{1.7E^{-6}}) * (\mathbf{10^6/T^2}) + (\mathbf{0.258} \pm \mathbf{1.7E^{-5}})$$

[Equation 1]

with an r^2 value of 0.93 and a *p*-value of <<0.0001 (**Figure 4c**). For comparison, the fit through the same data processed using Santrock/Gonfiantini parameters instead is:

$$\Delta_{47-RFAC(SG,P,newAFF)} = (\mathbf{0.0368} \pm \mathbf{1.7E^{-6}}) * (\mathbf{10^6/T^2}) + (\mathbf{0.280} \pm \mathbf{1.7E^{-5}})$$

[Equation 2]

with an r^2 value of 0.92 and a *p*-value of <<0.0001 (**Figure 4a**). Equations 1 and 2 are shown relative to a reaction at 25°C for historical reasons, but can be easily adjusted for different acid digestion temperatures (X) by adding back in the associated $\Delta^*_{25\text{-X}}$ value for $\Delta_{47\ (Br,P)}$ (see **Table 2**). The very low stated *p*-value rules out the null-hypothesis that temperature and Δ_{47} are uncorrelated, and does not directly speak to the goodness of fit.

Some datasets may have already been updated to IUPAC parameters prior to the publication of this study and therefore without access to our updated $\Delta_{47\text{-TE}}$ or $\Delta^*_{25\text{-X}}$ values (i.e. using WD $\Delta_{47\text{-TE}}$ values, and one of many possible $\Delta^*_{25\text{-X}}$ values). To facilitate more immediate use of this composite calibration, we also present fit parameters for the same synthetic dataset, using IUPAC parameters, WD $\Delta_{47\text{-TE}}$ values, and our new $\Delta^*_{25\text{-X}}$ values (**Table 3**), under the assumption that $\Delta^*_{25\text{-X}}$ values are simple to update.

Table 3 and **Table S4** include slopes, intercepts, and corresponding errors and measures of correlation strength for linear regressions through data processed different ways (SG vs. Br, WD vs. P) and for accompanying subsets of the data having formation temperatures <100°C. When selecting a calibration equation to use, it is vital that new unknown sample data be processed using the same parameters and $\Delta_{47\text{-TE}}$ values as the selected calibration equation, because although slopes are similar, intercepts can vary by ~0.05‰ (**Table 3, Table S4**). Small reported errors on calibration slope and intercept are the result of Monte-Carlo sampling through a very large dataset (*n*=451 synthetic replicates, *n*=379 synthetic <100°C, *n*=1253 total sample replicates, *n*=1181 total <100°C).

Overall, the magnitude of the improved inter-laboratory agreement obtained by updating to IUPAC parameters is limited (the minimal improvement in the composite synthetic calibration), indicating that remaining offsets must be caused by the continued presence of inter-laboratory inconsistencies related to other variables such as precipitation methods, CO_2 extraction techniques, other sample/standard preparation specifics, or choice of carbonates vs. gas standards for conversion to the absolute reference frame (Bernasconi et al., 2018) as opposed to raw data processing procedures, which were identical here.

3.5 Biogenic and natural inorganic samples similar to synthetic samples

Biogenic and natural inorganic samples included here (selected based on previously demonstrating apparent equilibrium behavior) behave similarly to synthetic carbonates (**Figure S5**). There is no visible trend in $\Delta\Delta_{47}$ (deviation from Δ_{47} predicted with **Equation 1 or 2**) which would indicate that biogenic carbonates follow a different calibration slope than synthetic carbonates (although we have not included corals, speleothems, or nautiloids). This synthetic-biogenic agreement has been seen in previous studies where both types of carbonates were analyzed in the same lab (e.g. Breitenbach et al., 2018). Additionally, the scatter in points around the composite calibration line (±0.05‰ at the replicate level, **Figure 4 and Figure S5**) is of the same magnitude for synthetic, biogenic, and natural inorganic samples, despite higher uncertainty in formation temperature for some biogenic and natural inorganic samples. Therefore, it is fair to extrapolate the calculated uncertainty and predictive power of the synthetic calibration (how well a temperature can be measured) to natural samples analyzed the same way.

Taking each study separately, the fit residuals decrease in three of the five studies (Petrizzo, Katz and Breitenbach, not Henkes or Wacker). When taken together, the conversion to IUPAC parameters does not noticeably improve the agreement between biogenic studies, but decreases the total residuals from the synthetic calibration line (calculated as the sum of the squares of all $\Delta\Delta_{47}$ values) slightly from 0.048 to 0.041 when IUPAC parameters are used (**Figure S5**).

Rather than representing a characteristic of biogenic samples in general, this minimal improvement is likely the result of the facts that 1) the labs in which biogenic calibration studies were performed tend to show smaller offsets between Santrock/Gonfiantini and IUPAC parameters than other labs (**Figure 3a**) and 2) biogenic samples tend to cover a smaller range in bulk composition than synthetically precipitated samples (**Figure S4**), providing less potential for large shifts. The spectacular discrepancies (up to 0.06‰) erased by updating to IUPAC parameters documented by Kelson et al. (2017) and Schauer et al. (2016) are only possible when the bulk composition of individual samples differ substantially, a feature of synthetically precipitated carbonates derived from different carbon sources/CO_2 tanks but rarely found in marine biogenic carbonates. However, we cannot rule out the possibility that the remaining discrepancies between these four biogenic studies (**Figure 7, Figure S5**) represent the existence of real differences between calibrations for various types of biogenic carbonates (foraminifera vs. mollusks vs. coccoliths, etc.).

3.6 No evidence for mineralogy-dependent calibration offsets

Theoretical modeling has suggested that different carbonate mineralogies should be defined by different, sub-parallel calibration lines (Δ_{47} vs. $10^6/T^2$), with offsets on the order of 0.03-0.05‰ (Schauble et al., 2006; Guo et al., 2009; Hill et al., 2014). These theoretical calibration lines combine predictions of equilibrium clumping in different carbonate mineralogies (Δ_{63}) (Schauble et al., 2006; Hill et al., 2014) with modeled mineralogical differences in the carbonate-to-CO_2 fractionation (e.g. $\Delta^*_{dolomite}$ or $\Delta^*_{calcite}$) (Guo et al., 2009). Contrary to these theoretical predictions, many, but not all (e.g. van Dijk et al., 2019), experimental studies have found that carbonates of different mineralogies measured in the same laboratory seem to follow identical calibration lines (Bonifacie et al., 2017; Breitenbach et al., 2018; Defliese et al., 2015; Kluge and John, 2015; Winkelstern et al., 2016). We examine if offsets between carbonate mineralogies exist in the large and mineralogically-varied dataset compiled for this study.

Figure 5 compares calcite and non-calcite mineralogies corrected using both parameter sets. Updating to IUPAC parameters erases differences on the order of 0.025‰ between dolomite and calcite data in our dataset (**Figure 5**). This is likely not a mineralogical effect, but instead due to the fact that all dolomite data here was measured at University of Michigan, which shows the largest effects of updating to IUPAC parameters (**Figure 3a**). In addition to being apparent in the overlapping 95% confidence intervals shown in **Figure 5b** where each mineralogy is considered separately (as a linear fit through offsets from a calcite-only regression), this lack of distinction between mineralogies also holds up statistically. An ANCOVA analysis results in slopes and intercepts that are not statistically differentiable between all pairs of mineralogies when using IUPAC parameters. Only dolomite has a statistically distinguishable intercept from other mineralogies (all excluding magnesite) and slope (compared to calcite only) when using Santrock/Gonfiantini parameters, as is also indicated in the clear offset between dolomite and the other mineralogies that can be seen in **Figure 5a**.

Bonifacie et al. (2017) observed $\Delta^*_{dolomite90}$ values (difference between Δ_{47} of CO_2 extracted in 90°C acid and theoretical Δ_{63} of solid dolomite) of +0.176‰, lower than the +0.198‰ $\Delta^*_{calcite90}$ values seen by Passey and Henkes (2012) (see Bonifacie et al. (2017) Figure 4). Use of IUPAC parameters reduces Passey $\Delta^*_{calcite90}$ values slightly and also brings high-temperature dolomite data from Winkelstern et al. (2016) into better alignment with other dolomite data (Bonifacie et al. 2017), which although not updated here, is not expected to change much based on changes in another dataset measured in the same lab (Katz et al., 2017) (**Figure S6**).

The fact that calcite, aragonite, and dolomite (at the least) appear to follow identical calibration lines in this dataset does not necessarily mean that mineralogical differences in clumping do not exist within the carbonate. It is possible that equilibrium clumping varies between carbonate mineralogies (in Δ_{63}), but mineralogical differences in the full acid fractionation factor $\Delta^*_{mineralogy}$ vary such that they cancel out (within the error of our measurements), resulting in a single apparent relationship between temperature and Δ_{47} that can be applied to all six included mineralogies. However, theory predicts the opposite - that mineralogical differences in Δ_{63} clumping and in the carbonate-to-CO_2 fractionation ($\Delta^*_{mineralogy}$) are additive, resulting in offsets of ~0.05‰ between calibration lines (e.g. Guo et al., 2009).

3.7 Assessment of the coherence of the composite calibration equation

One issue that has existed in the clumped isotope literature is the “two slope problem” evident in the large disagreement between the first two clumped isotope calibration studies published (Dennis and Schrag, 2010; Ghosh et al., 2006). Even after retroactive updating to the absolute reference frame, these two calibrations are offset by the equivalent of ~14°C around 0°C and by ~30°C around 100°C, crossing in between (Dennis et al., 2011). This expanded into a "multi slope problem" with the publication of a travertine calibration by Kele et al. (2015), with a slope intermediate between the first two studies (Kele = 0.044 vs. Dennis = 0.036 vs. Ghosh = 0.063 x 10^6), and most subsequent studies have fallen somewhere within this range. Direct comparison of intercepts in subsequent studies has been more difficult due to use of evolving acid digestion corrections over time. We can use our multi-study composite dataset to investigate the range of slopes and intercepts in individual studies and subsets of the full dataset, when all data is processed identically and uses the same $\Delta^*_{25\text{-}X}$ value set across all studies.

Fit residuals *within* individual studies are improved in 11 out of 14 cases (excluding Passey, Henkes, and Wacker), as evidenced by increased r^2 values in linear regressions through data from each individual study (**Figure S7, Table S5**). As an example, the Kelson data was previously shown to collapse to a much tighter cluster when using IUPAC parameters, as seen in Figures 4 and 5 of the original publication (Kelson et al., 2017). Taking all 13 low-temperature studies individually (i.e. excluding Passey), the range in Δ_{47} values predicted for a temperature of 25°C decreases from 0.060‰ to 0.037‰ and the range in temperature predicted for a Δ_{47} value of 0.700‰ decreases from 20°C to 12°C with the adoption of IUPAC parameters (Figure S7).

None of the studies included here produce slopes close to 0.06 x 10^6 (i.e. similar to Ghosh et al. (2006)). In fact, no study with a slope greater than 0.046 has been produced since the adoption of the absolute reference frame, and thus none met our criteria for inclusion. For both the SG and IUPAC parameters, we see a convergence of slopes into the range of 0.035-0.038 for different subsets of the multi-study composite dataset (**Table 3**, **Table S4**), and into the somewhat wider range of 0.033-0.041 for individual studies (albeit with larger errors due to smaller number of samples per fit) (**Figure 6, Table S5, Figure S7**). This convergence in slope is reassuring, in that it appears that the community has resolved the original "two slope" problem (Bonifacie et al., 2017; Fernandez et al., 2017; Kelson et al., 2017), but smaller-scale disagreement between studies remains. **Figure 7** shows the offsets of individual studies from the synthetic composite calibration described in **Equation 1**. It is probable that remaining offsets are not solely the result of noise (or random error) in the measurement – individual labs and studies show coherent offsets from the mean (e.g. all samples from Katz fall above the composite calibration) and internal scatter within one study is usually less than the full range of scatter amongst studies (e.g. Peral, Wacker, Fernandez). This suggests that current standardization protocols may not successfully eliminate inter-laboratory biases, causing inter-laboratory differences to persist. Additionally, this implies that temperature estimates from a calibration derived within the same lab as unknowns may be more accurate than a composite calibration combining data from many labs.

Differences in the scatter in calibration data between studies may reflect a sampling strategy rather than anything inherent in the laboratory practices or nature and homogeneity of the samples. Some studies (Tang et al., 2014; Kelson et al., 2017) chose to replicate experiments at the carbonate precipitation stage, making many independent aliquots of synthetic carbonate at a given formation temperature, but measuring each one only a few times (N = ~3). Other studies chose to focus on fewer samples, but measured each one more times each (Breitenbach et al., 2018), reducing the uncertainty on a given sample mean value. Such differences in sampling strategy, and previously noted differences in the total number and temperature range of data points included in an individual published calibration (e.g. Fernandez et al., 2017; Kelson et al., 2017), impact calibration discrepancies as well as uncertainties in calculated temperature for sample unknowns.

Although it is mathematically straightforward to aggregate observations from all calibration studies and compute an overall regression model for the whole dataset (which we have done in **Equation 1** for all synthetic data and report in **Table S4** for other subsets of data), doing so is only meaningful if we have confidence that all studies reflect a unique underlying "true" relationship between Δ_{47} and temperature, with random observation errors. **Figure 6 and Table S6** summarize the results of the ANCOVA tests investigating whether differences in T-Δ_{47} relationships between labs are statistically significant, using the most updated dataset ($\Delta_{47\text{-RFAC(Br,P,newAFF)}}$). Out of the 91 calibration pairs considered here, only 10 yield statistically significant differences in slope (at the 95% confidence level), which is reassuring. The remaining 81 pairs yield statistically indistinguishable slopes (42 of which are also statistically indistinguishable in intercept). These numbers indicate pairwise agreement. Considering all mathematically possible values of slope and intercept, at most 9 studies out of 14 can be in simultaneous agreement. These results only correspond to a slight improvement over the use of the Gonfiantini et al. (1995) ^{17}O correction parameters in the pairwise comparisons (same number of pairs indistinguishable for slope/intercept; 10 with differences in both slope/intercept vs. 11 for SG/P). Synthetic studies in particular are brought into closer agreement with use of IUPAC parameters (Kelson, Kluge, Winkelstern). Whether the results of this comparison exercise are interpreted as encouraging or not, it appears clear that the use of new ^{17}O correction parameters does not fully reconcile the results from different calibration studies.

It is possible that remaining differences between calibration studies result from systematic analytical biases between laboratories, use of carbonate vs. gas standards to convert into the absolute reference frame, the existence of different T-Δ_{47} relationships in different types of carbonates (e.g. Daëron et al., 2019), precipitation conditions of synthetic carbonates (e.g. pH, saturation, precipitation rate), or other details of sample preparation. If so, it becomes problematic to justify the use of a "universal" regression line obtained from the aggregated dataset. However, the convergence of slope in many of these studies suggests we are getting closer to understanding the underlying nature of the T-Δ_{47} relationship.

3.8 Causes of remaining inter-laboratory differences

Although differences between calibrations are increasingly minor, their persistence suggests that further improvement in clumped isotope methods and calibration is necessary. Here, we have eliminated a number of possible causes of inter-laboratory disagreement. By restricting ourselves to studies performed after the adoption of the absolute reference frame, where the gas and/or carbonate standards analyzed concurrently with unknowns were selected with the purpose of tracking changes in the reference frame, we rule out a poorly-constrained gas-based reference frame as a source of uncertainty (although systematic study of the errors associated with conversion to the absolute reference frame has never been carried out, and some suggest that gas-based reference frames may be inadequate (Bernasconi et al., 2018)). We have also eliminated variations in the temperature dependence of the acid digestion fractionation ($\Delta^*_{25\text{-X}}$), $\Delta_{47\text{-TE}}$ values, or ^{17}O correction parameters (which were not uniformly applied across studies in the original publications – e.g. Petrizzo et al., 2014) as causes of the remaining disagreement. Obvious calibration differences due to methods of sample cleaning (GC vs. He-carrier PPQ vs. static PPQ trap), mass spectrometer (Thermo 253 or 253+ vs. Isoprime 100 vs. Thermo 253 with peak hopping), or background/pressure baseline measurement method (once per day or longer vs. once per sample vs. once per acquisition vs. none at all) are not apparent from comparisons of data generated in different laboratories. However, other differences are less easily assessed, even via the extensive data reprocessing exercise presented here. As suggested by several studies (Bajnai et al., 2018; Daëron et al., 2019; Davies and John, 2018; Dennis et al., 2013; Saenger et al., 2012, 2017; Spooner et al, 2016), overall and internal isotopic equilibrium appears to not always be attained during biologically mediated carbonate precipitation, and slight and variable kinetically-induced departures from equilibrium might also occur during laboratory precipitation of synthetic carbonates. Moreover, Bernasconi et al. (2018) and Bonifacie et al. (2017) pointed out that steep and rapidly changing mass spectrometer linearity may contribute to uncertainty and discrepancies among calibration datasets. Reference frame stability and correction window choice were not investigated here, and practices vary dramatically between laboratories based on the behavior of their individual mass spectrometer. Additionally, background/pressure baseline correction and data reduction methods should be further investigated. While moving to a carbonate standard-based correction scheme would not necessarily reveal which of these slight differences in method is causing the remaining minor disagreements, their use has the promising potential to remove persistent inter-lab biases in both calibration and acid digestion fractionation data (Bernasconi et al, 2018; Peral et al., 2018).

Our results support the idea that current best practices in data calibration and reduction do noticeably (and in some cases dramatically) improve the fidelity of Δ_{47} data; yet it seems that a truly universal calibration remains elusive. Going forward, laboratories may reasonably choose to continue to use their in-house generated calibrations to account for artifacts by lab- or instrument-specific sample preparation and analytical conditions, particularly if the in-house calibration is based on many samples spanning a large temperature range bracketing sample unknowns. In other cases, it may be beneficial (ie. in terms of confidence in and precision of temperature estimates) to take advantage of the large sample size and temperature range represented by the composite calibration presented in **Equation 1**. Workers could also choose to “pin” themselves to this or another existing calibration via the analysis of carbonate standards and/or synthetic carbonates, but use the slope defined by **Equation 1**.

3.9 Performance of the clumped isotope paleothermometer

Researchers both within and outside of the clumped isotope community may ask "how good is the clumped isotope paleothermometer?" or "How big is the error on a typical clumped isotope temperature?". Towards this end, we lay out the following observations.

Typical long-term reproducibility (1sd) on the Δ_{47} measurement is ~0.02‰ (or ~0.04‰ for the Kiel microsampling method where unknowns are replicated many more times) (**Table 1**). Assuming N replicates per unknown (typically 3 is the minimum number of replicates considered "best practice" for labs using traditional sample sizes of 3-8mg/replicate), this translates to an external standard error of $0.02/\sqrt{N}$ on $\Delta_{47\text{-RFAC}}$ (0.012‰ for N=3). Depending on the sample mean Δ_{47} value, the same level of uncertainty in Δ_{47} translates to a vastly different amount of uncertainty in temperature due to the shape of the relationship between temperature and Δ_{47}. Colored lines in **Figure 8** shows how the error in temperature grows as temperature increases (and Δ_{47} decreases) for N=1, 4, and 10.

Uncertainty in the slope and intercept of the T-Δ_{47} calibration equation also contribute to total uncertainty on a clumped isotope temperature. Although the reported errors on the slope and intercept in **Equation 1** are quite small due to the Monte Carlo sampling procedure and the large number of samples included, this overestimates how well we actually know the true slope and intercept of the T-Δ_{47} calibration (not to mention that the existence of a single universal calibration is still debated). The persistence of differences between single-lab calibrations (as demonstrated by offsets in **Figure 7** and statistically distinguishable slopes and/or intercepts in 49 pairwise comparisons shown in **Table S6**) indicates that existing and new labs still need to constrain the T-Δ_{47} calibration internally in order to produce robust clumped isotope temperatures. It appears that the remaining differences are largely in the intercept as opposed to the slope. The convergence of calibration slopes shown in **Figure 6, Table S5,** and **Table S6** mean the community is already able to robustly reconstruct relative temperature change across many labs. Despite remaining inter-laboratory differences elevating the apparent uncertainty in the T-Δ_{47} calibration, within a single lab using an internally well-constrained calibration (including N>30 replicates), the analytical error still dwarfs the error contributed by the uncertainty in slope and intercept (black lines, **Figure 8**). As a result, total uncertainty on a typical clumped isotope temperature is dominated by analytical error.

4. Future Recommendations for Δ_{47} Data Reduction and Reporting

This compilation study has brought to light some room for improvement in data processing and reporting within the clumped isotope community and moves us to make some recommendations for future studies.

To begin with, we recommend using the updated IUPAC parameters for all data processing from this point forward. For best accuracy, established values for all internal laboratory carbonate standards should also be updated or re-established using IUPAC parameters. We recommend using a consistent set of $\Delta_{47\text{-TE}}$ values (listed in the read-off table in **Table S2**) instead of the polynomial fit equation in Dennis et al. (2011) to calculate the empirical transfer function. We recommend using the new $\Delta^*_{25\text{-X}}$ values determined from four combined, reprocessed studies (**Table 2**) to correct for the fractionating effects of acid digestion.

To calculate temperatures for sample unknowns, it is important to use a calibration equation based on data processed using the same parameters as the unknowns. This includes the choice of parameters (IUPAC vs. Santrock/Gonfiantini), $\Delta_{47\text{-TE}}$ values and acid fractionation $\Delta^*_{25\text{-X}}$ values.

While there is no easy correction that can be applied to published Δ_{47} data calculated with S/G parameters, workers may be able to predict the magnitude of change that could be expected by looking at the bulk isotopic composition of sample and standards and the magnitude of change ($\Delta\Delta_{47\text{-RF}}$ (SG-Br)) calculated here for various studies. For example, although a portion of the dolomite data in Bonifacie et al. (2017) was measured at Caltech, the fact that other data measured in the same lab (IPGP, Katz et al., 2017) did not change much suggests that the IPGP portion of the Bonifacie et al. (2017) data also would not change much.

It has already become general practice to include data tables of replicate-level sample data as material accompanying published works. We support this and advocate that it become universally adopted. In addition to being good for scientific transparency and data accessibility within the community, it aligns with new open-access policies put forward by funding agencies and publishers. However, we found that existing tables rarely contained the sufficiently raw data (replicate mean δ^{45}-δ^{49} values, post-background-correction) needed to reprocess using updated parameters such as was done in this study. It is unlikely that the IUPAC parameter set is the final say on the true values of R^{13}_{VPDB}, R^{17}_{VSMOW}, R^{18}_{VSMOW} and λ, and these parameters may be updated again in the future, requiring another round of reprocessing to utilize older published clumped isotope data. Therefore, we recommend that all future published studies include **1)** replicate mean δ^{45}-δ^{49} values for <u>all</u> sample and standard replicates (in-house or community-wide carbonates, equilibrated or heated gases), **2)** sufficient information about how ARF parameters were calculated (which standards were used to correct which samples, correction intervals, moving windows, etc.), as well as **3)** reference (working) gas composition at the time of analysis. This is the minimum needed to perform the data reprocessing described in this study. Software platforms like Easotope (John and Bowen, 2016) make exporting this level of data to Excel very simple, and can go further by exporting a full dataset for a particular study with the "trim database" function. Having a dataset available as a mini-Easotope database allows external users to reprocess the data at will and supports end-to-end transparency in data selection and processing. Reporting of carbonate standard data (whether used for reference frame calculations or not) is highly encouraged and will aid in inter-study comparison as well as documentation of long-term, external reproducibility. Additionally, information about preparation techniques like reaction temperature, reaction vessel type, cleaning method, etc., and information about the samples themselves like mineralogy, collection locality, carbonate type/taxa, and geologic age may be vital in future reprocessing efforts and should also be included.

To facilitate improved and standardized data reporting, we have established a data template with the archiving database EarthChem that includes columns for the necessary raw data, as well as sample collection and preparation information, and provides a doi number for each archived dataset. All datasets in this study will be archived in this database along with publication of this manuscript, and a blank template will be available from EarthChem for future use.

5. Summary and Conclusions

The clumped isotopic composition of CO_2 derived from carbonate (denoted Δ_{47}) is a function of carbonate formation temperature, and in natural samples can act as a recorder of paleoclimate, burial or diagenetic conditions, depending on the sample type. In order to calculate Δ_{47}, values defining the absolute abundance of heavy isotopes in the universal standards VPDB and VSMOW (defined by four parameters: R^{13}_{VPDB}, R^{17}_{VSMOW}, R^{18}_{VSMOW} and λ) are incorporated into calculations at a very early level. These values have been recently revised and updated from the “Santrock/Gonfiantini” parameter set defined in early clumped isotope publications (Huntington et al., 2009) to newer values defined by Brand et al. (2010). The purpose of this study was to determine the effects of this parameter update, and to determine whether this could resolve ongoing inter-laboratory discrepancies in calibration data.

We assembled and reprocessed data from 14 studies (8 synthetic, 6 biogenic/natural), including carbonates of variable mineralogy, bulk composition, and formation temperature, measured in 10 different laboratories using different reaction temperatures, cleaning procedures, and mass spectrometers. Beginning from either raw voltages or sample mean δ^{45} to δ^{49} values, data was reprocessed two ways, using the “Santrock/Gonfiantini” parameters defined by Huntington et al. (2009) or the updated “IUPAC” parameters defined by Brand et al. (2010).

In 11 out of 14 studies, data processed using IUPAC parameters showed improved agreement (determined by the r^2 value of a linear fit through data from each individual study) compared to the same data processed using Santrock/Gonfiantini parameters. A composite calibration combining all 8 synthetic studies (**Equation 1**) showed only marginally improved statistics, although scatter between samples with high formation temperatures visibly diminished. The magnitude of the change in $\Delta_{47\text{-RFAC}}$ produced by updating to IUPAC parameters is complex and dependent on both the bulk composition of an individual sample and the bulk composition of the gas or carbonate standards used to correct that sample into the absolute reference frame. As a result, no “transfer function” can be used to easily update old data and data must be reprocessed from the rawest format.

This large compilation of clumped isotope data including data measured many different ways but well-referenced to standards and processed identically gives us the unique ability to test whether a universal calibration may exist for carbonate clumped isotopes. From this dataset, there is no indication that carbonates of different mineralogy or sample type (synthetic vs. natural or biogenic) require different calibration equations. Acid digestion fractionation data of all mineralogies agree as well, although no studies initially reporting larger (order +0.15‰) fractionation factors were included and reprocessed here. There is no observed difference in data from labs using or not using the pressure baseline (PBL) correction. Type of mass spectrometer and method of sample cleaning (GC vs. PPQ trap) also do not appear to cause systematic offsets.

Slopes of individual calibrations are clustered around 0.033-0.041 x 10^6 (excluding Passey, which has a very narrow Δ_{47} range) and 81 out of 91 pairwise comparisons show indistinguishable slopes at the 95% confidence level, demonstrating that the “multiple slopes” issue that plagued the clumped isotope community in its infancy has been largely eliminated (although no studies initially showing the steeper slopes met all criteria for inclusion in this study, so direct comparison was impossible). Additionally, the range in Δ_{47} values predicted for 25°C from 14 individual calibrations decreases from 0.060‰ to 0.038‰ with the adoption of IUPAC parameters and 9 out of 14 studies can be in simultaneous agreement in slope and intercept within error, showing increased coherence across the community. Whether you interpret this improved agreement as encouraging or view the remaining discrepancies as discouraging, this study shows that the continued differences in intercept (equivalent to ± 6°C) must be due to other differences in analytical conditions that remain to be systematically investigated. Such investigations may be enabled by increased use of internationally available carbonate standards (Bernasconi et al., 2018) in addition to using identical parameter sets and constants presented here.

In this study, we have ruled out several possible causes of these remaining discrepancies – compositionally dependent biases introduced by the use of incorrect standard parameters, data reduction methods, and use of different acid fractionation corrections and $\Delta_{47\text{-TE}}$ values. This still leaves many future avenues for investigation and improvement.

Acknowledgements and Data Statement

Originally processed data is included in cited papers. Reprocessed samples, replicates, and means are included in this manuscript as supplementary data tables. All reprocessed datasets are also archived in the EarthChem database using a new data template designed specifically for carbonate clumped isotope data as part of this study. Each reprocessed study has been archived individually and assigned its own doi number as follows: Winkelstern et al., 2016 (10.1594/IEDA/111317); Defliese et al., 2015 (10.1594/IEDA/111318); Kluge et al., 2015 (10.1594/IEDA/111319); Garcia de Real et al., 2016 (10.1594/IEDA/111320); Kelson et al., 2017 (10.1594/IEDA/111321); Tang et al., 2014 (10.1594/IEDA/111322); Fernandez et al., 2014 (10.1594/IEDA/111323); Passey and Henkes, 2012 (10.1594/IEDA/111324); Henkes et al., 2013 (10.1594/IEDA/111325); Peral et al., 2018 (10.1594/IEDA/111326); Wacker et al., 2014

(10.1594/IEDA/111327); Petrizzo et al., 2014 (10.1594/IEDA/111328); Katz et al., 2017 (10.1594/IEDA/111329); Breitenbach et al., 2018 (10.1594/IEDA/111330).

The authors acknowledge J. Tang, P. Garcia de Real, and U. Wacker for producing and publishing some of the data included in this study. The original Kluge et al. (2015) calibration (CJ, TK) was part of the Qatar Carbonate and Carbon Sequestration Research Center, funded by Qatar Petroleum, Shell, and the Qatar Science and Technology Park. DAH and SFMB were supported by NERC project ID NE/M003752/1. SK was supported by the National Research, Development and Innovation Office NKF1H, KH 125584. KWH acknowledges US National Science Foundation grants EAR-1156134 and 1713275. This work was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division under Award Number DE-SC0010288 to AT. This is IPGP contribution 3995.

Author contribution statement

The project to reprocess raw data from published calibration studies using IUPAC parameters was initiated independently by several groups including (in no particular order) coauthors SVP, AJS, KWH, MD, and ASC. These efforts were combined into one project spearheaded by SVP, who designed the present study, reprocessed all data, wrote manuscript. WFD performed $\Delta_{47\text{-TE}}$ calculations. CS performed Monte Carlo analysis. MD performed ANCOVA analysis and error propagation in reconstructed temperatures and made figures. JRK compiled references. CS and SVP designed EarthChem template. CS, MD, CMJ, SMB, ASC, KWH, JRK, TK, GAO, AJS consulted on study design. For each published study providing data, the first author and laboratory lead were included as coauthors, plus additional workers not involved in the original publications but who contributed significantly to assembling data for reprocessing. These "data contributing authors" are listed in alphabetical order, but are attributed to each study as follows: Defliese (WFD, KCL), Winkelstern (IZW, KCL), Kluge (TK, CMJ), Garcia (CMJ), Kelson (JRK, KWH), Tang (RV, BER, AT), Fernandez (ABF, RV, BER), Passey (BHP), Peral (MD, MP), Henkes (GAH, BHP), Petrizzo (DAP, EDY), Wacker (UW, DB, JF), Katz (AK, MB), Breitenbach (SFMB, DAH). All authors edited and approved manuscript.

References

Affek, H.P., Bar-Matthews, M., Ayalon, A., Matthews, A., and Eiler, J.M. (2008) Glacial/interglacial temperature variations in Soreq cave speleothems as recorded by 'clumped isotope' thermometry. *Geochimica et Cosmochimica Acta*, *72*, 5351-5360. https://doi.org/10.1016/J.GCA.2008.06.031

Baertschi, P. (1976). Absolute 18O content of standard mean ocean water. *Earth and Planetary Science Letters*, *31*(3), 341–344. https://doi.org/10.1016/0012-821X(76)90115-1

Bajnai, D., Fiebig, J., Tomašových, A., Milner Garcia, S., Rollion-Bard, C., Raddatz, J., et al. (2018). Assessing kinetic fractionation in brachiopod calcite using clumped isotopes. *Scientific Reports*, *8*(1), 533. https://doi.org/10.1038/s41598-017-17353-7

Bernasconi, S. M., Hu, B., Wacker, U., Fiebig, J., Breitenbach, S. F. M., & Rutz, T. (2013). Background effects on Faraday collectors in gas-source mass spectrometry and implications for clumped isotope measurements. *Rapid Communications in Mass Spectrometry*, *27*(5), 603–612. https://doi.org/10.1002/rcm.6490

Bernasconi, S. M., Müller, I. A., Bergmann, K. D., Breitenbach, S. F. M., Fernandez, A., Hodell, D. A., et al. (2018). Reducing uncertainties in carbonate clumped isotope analysis through consistent carbonate-based standardization. *Geochemistry, Geophysics, Geosystems*, 1–20. https://doi.org/10.1029/2017GC007385

Bonifacie, M., Calmels, D., Eiler, J. M., Horita, J., Chaduteau, C., Vasconcelos, C., et al. (2017). Calibration of the dolomite clumped isotope thermometer from 25 to 350 °C, and implications for a universal calibration for all (Ca, Mg, Fe)CO3 carbonates. *Geochimica et Cosmochimica Acta*, *200*, 255–279. https://doi.org/10.1016/J.GCA.2016.11.028

Brand, W. A., Assonov, S. S., & Coplen, T. B. (2010). Correction for the 17O interference in δ(13C) measurements when analyzing CO2 with stable isotope mass spectrometry (IUPAC Technical Report). *Pure and Applied Chemistry*, *82*(8), 1719–1733. https://doi.org/10.1351/PAC-REP-09-01-05

Breitenbach, S. F. M., Mleneck-Vautravers, M. J., Grauel, A.-L., Lo, L., Bernasconi, S. M., Müller, I. A., et al. (2018). Coupled Mg/Ca and clumped isotope analyses of foraminifera provide consistent water temperatures. *Geochimica et Cosmochimica Acta*, *236*, 283–296. https://doi.org/10.1016/J.GCA.2018.03.010

Burgener, L., Huntington, K. W., Hoke, G. D., Schauer, A., Ringham, M. C., Latorre, C., & Díaz, F. P. (2016). Variations in soil carbonate formation and seasonal bias over >4 km of relief in the western Andes (30°S) revealed by clumped isotope thermometry. *Earth and Planetary Science Letters*, *441*, 188–199. https://doi.org/10.1016/j.epsl.2016.02.033

Came, R. E., Brand, U., & Affek, H. P. (2014). Clumped isotope signatures in modern brachiopod carbonate. *Chemical Geology*, *377*, 20–30. https://doi.org/10.1016/j.chemgeo.2014.04.004

Came, R. E., Azmy, K., Tripati, A., & Olanipekun, B.-J. (2017). Comparison of clumped isotope signatures of dolomite cements to fluid inclusion thermometry in the temperature range of 73–176 °C. *Geochimica et Cosmochimica Acta*, *199*, 31–47. https://doi.org/10.1016/J.GCA.2016.10.028

Came, R. E., Eiler, J. M., Veizer, J., Azmy, K., Brand, U., & Weidman, C. R. (2007). Coupling of surface temperatures and atmospheric CO2 concentrations during the Palaeozoic era. *Nature*, *449*, 198. http://dx.doi.org/10.1038/nature06085

Cao, X. & Liu, Y., 2012. Theoretical estimation of the equilibrium distribution of clumped isotopes in nature. *Geochimica et Cosmochimica Acta*, *77*, 292–303. https://doiorg/10.1016/j.gca.2011.11.021

Craig, H. (1957). Isotopic standards for carbon and oxygen and correction factors for mass-spectrometric analysis of carbon dioxide. *Geochimica et Cosmochimica Acta*, *12*(1–2), 133–149. https://doi.org/10.1016/0016-7037(57)90024-8

Daëron, M., Blamart, D., Peral, M., & Affek, H. P. (2016). Absolute isotopic abundance ratios and the accuracy of Δ47 measurements. *Chemical Geology*, *442*, 83–96. https://doi.org/10.1016/j.chemgeo.2016.08.014

Daëron, M., Drysdale, R.N., Peral, M., Huyghe, D., Blamart, D., Coplen, T.B., et al. (2019) Most Earth-surface calcites precipitate out of isotopic equilibrium. *Nature Communications*, *10*:*429*, 1-7. https://doi.org/10.1038/s41467-019-08336-5

Daëron, M., Guo, W., Eiler, J., Genty, D., Blamart, D., Boch, R., et al. (2011). 13C18O clumping in speleothems: Observations from natural caves and precipitation experiments. *Geochimica et Cosmochimica Acta*, *75*(12), 3303–3317. https://doi.org/10.1016/j.gca.2010.10.032

Davies, A.J. & John, C.M. (2019). The clumped (13C18O) isotope composition of echinoid calcite: Further evidence for “vital effects” in the clumped isotope proxy. *Geochimica et Cosmochimica Acta*, *245*, 172–189. https://doi.org/10.1016/j.gca.2018.07.038

Defliese, W. F., Hren, M. T., & Lohmann, K. C. (2015). Compositional and temperature effects of phosphoric acid fractionation on Δ47 analysis and implications for discrepant calibrations. *Chemical Geology*, *396*, 51–60. https://doi.org/10.1016/j.chemgeo.2014.12.018

Dennis, K. J., Affek, H. P., Passey, B. H., Schrag, D. P., & Eiler, J. M. (2011). Defining an absolute reference frame for ‘clumped’ isotope studies of CO2. *Geochimica et Cosmochimica Acta*, *75*(22), 7117–7131. https://doi.org/10.1016/j.gca.2011.09.025

Dennis, K. J., & Schrag, D. P. (2010). Clumped isotope thermometry of carbonatites as an indicator of diagenetic alteration. *Geochimica et Cosmochimica Acta*, *74*(14), 4110–4122. https://doi.org/10.1016/j.gca.2010.04.005

Dennis, K. J., Cochran, J. K., Landman, N. H., & Schrag, D. P. (2013). The climate of the Late Cretaceous: New insights from the application of the carbonate clumped isotope thermometer to Western Interior Seaway macrofossil. *Earth and Planetary Science Letters*, *362*, 51–65. https://doi.org/10.1016/j.epsl.2012.11.036

Douglas, P. M. J., Affek, H. P., Ivany, L. C., Houben, A. J. P., Sijp, W. P., Sluijs, A., et al. (2014). Pronounced zonal heterogeneity in Eocene southern high-latitude sea surface temperatures. *Proceedings of the National Academy of Sciences*, *111*(18), 6582–6587. https://doi.org/10.1073/pnas.1321441111

Eagle, R. A., Schauble, E. A., Tripati, A. K., Tütken, T., Hulbert, R. C., & Eiler, J. M. (2010). Body temperatures of modern and extinct vertebrates from (13)C-(18)O bond abundances in bioapatite. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(23), 10377–10382. https://doi.org/10.1073/pnas.0911115107

Eagle, R. A., Enriquez, M., Grellet-Tinner, G., Pérez-Huerta, A., Hu, D., Tütken, T., et al. (2015). Isotopic ordering in eggshells reflects body temperatures and suggests differing thermophysiology in two Cretaceous dinosaurs. *Nature Communications*, *6*, 8296. https://doi.org/10.1038/ncomms9296

Eagle, R. A., Eiler, J. M., Tripati, A. K., Ries, J. B., Freitas, P. S., Hiebenthal, C., et al. (2013). The influence of temperature and seawater carbonate saturation state on 13C–18O bond ordering in bivalve mollusks. *Biogeosciences*, *10*(7), 4591–4606. https://doi.org/10.5194/bg-10-4591-2013

Eiler, J. M. & Schauble, E. (2004). 18O13C16O in Earth's atmosphere. *Geochimica et Cosmochimica Acta*, *68*, 4767-4777. https://doi.org/10.1016/j.gca.2004.05.035

Falk, E. S., & Kelemen, P. B. (2015). Geochemistry and petrology of listvenite in the Samail ophiolite, Sultanate of Oman: Complete carbonation of peridotite during ophiolite emplacement. *Geochimica et Cosmochimica Acta*, *160*, 70–90. https://doi.org/10.1016/J.GCA.2015.03.014

Fernandez, A., Müller, I. A., Rodríguez-Sanz, L., van Dijk, J., Looser, N., & Bernasconi, S. M. (2017). A Reassessment of the Precision of Carbonate Clumped Isotope Measurements: Implications for Calibrations and Paleoclimate Reconstructions. *Geochemistry, Geophysics, Geosystems*, 1–12. https://doi.org/10.1002/2017GC007106

Fernandez, A., Tang, J., & Rosenheim, B. E. (2014). Siderite 'clumped' isotope thermometry: A new paleoclimate proxy for humid continental environments. *Geochimica et Cosmochimica Acta*, *126*, 411–421. https://doi.org/10.1016/j.gca.2013.11.006

Fiebig, J., Hofmann, S., Löffler, N., Lüdecke, T., Methner, K., & Wacker, U. (2016). Slight pressure imbalances can affect accuracy and precision of dual inlet-based clumped isotope analysis. *Isotopes in Environmental and Health Studies*, *6016*(April), 1–17. https://doi.org/10.1080/10256016.2015.1010531

Fisher, J. & Ullrich, J. (2016). THe new system of units. *Nature Physics*, *12*, 4-7.

García del Real, P., Maher, K., Kluge, T., Bird, D. K., Brown, G. E., & John, C. M. (2016). Clumped-isotope thermometry of magnesium carbonates in ultramafic rocks. *Geochimica et Cosmochimica Acta*, *193*, 222–250. https://doi.org/10.1016/J.GCA.2016.08.003

Ghosh, P., Adkins, J., Affek, H., Balta, B., Guo, W., Schauble, E. A., et al. (2006). 13C–18O bonds in carbonate minerals: A new kind of paleothermometer. *Geochimica et Cosmochimica Acta*, *70*(6), 1439–1456. https://doi.org/10.1016/j.gca.2005.11.014

Ghosh, P., Eiler, J., Campana, S. E., & Feeney, R. F. (2007). Calibration of the carbonate 'clumped isotope' paleothermometer for otoliths. *Geochimica et Cosmochimica Acta*, *71*(11), 2736–2744. https://doi.org/10.1016/j.gca.2007.03.015

Gonfiantini, R., Stichler, W., & Rozanski, K. (1995). *Standards and Intercomparison Materials Distributed by the International Atomic Energy Agency for Stable Isotope Measurements*. https://doi.org/10.1016/0375-6505(95)00024-0

Grauel, A.-L., Schmid, T. W., Hu, B., Bergami, C., Capotondi, L., Zhou, L., & Bernasconi, S. M. (2013). Calibration and application of the 'clumped isotope' thermometer to foraminifera for high-resolution climate reconstructions. *Geochimica et Cosmochimica Acta*, *108*, 125–140. https://doi.org/10.1016/j.gca.2012.12.049

Guo, W., Mosenfelder, J. L., Goddard, W. A., & Eiler, J. M. (2009). Isotopic fractionations associated with phosphoric acid digestion of carbonate minerals: Insights from first-principles theoretical modeling and clumped isotope measurements. *Geochimica et Cosmochimica Acta*, *73*(24), 7203–7225. https://doi.org/10.1016/j.gca.2009.05.071

He, B., Olack, G. A., & Colman, A. S. (2012). Pressure baseline correction and high-precision CO 2 clumped-isotope (Δ 47) measurements in bellows and micro-volume modes. *Rapid Communications in Mass Spectrometry*, *26*(24), 2837–2853. https://doi.org/10.1002/rcm.6436

Henkes, G. A., Passey, B. H., Wanamaker, A. D., Grossman, E. L., Ambrose, W. G., & Carroll, M. L. (2013). Carbonate clumped isotope compositions of modern marine mollusk and brachiopod shells. *Geochimica et Cosmochimica Acta*, *106*, 307–325. https://doi.org/10.1016/j.gca.2012.12.020

Hill, P. S., Tripati, A. K., & Schauble, E. a. (2014). Theoretical constraints on the effects of pH, salinity, and temperature on clumped isotope signatures of dissolved inorganic carbon species and precipitating carbonate minerals. *Geochimica et Cosmochimica Acta*, *125*, 610–652. https://doi.org/10.1016/j.gca.2013.06.018

Huntington, K. W., Eiler, J. M., Affek, H. P., Guo, W., Bonifacie, M., Yeung, L. Y., et al. (2009). Methods and limitations of “clumped” CO2 isotope (Δ47) analysis by gas-source isotope ratiomass spectrometry. *Journal of Mass Spectrometry*, *44*(9), 1318–1329. https://doi.org/10.1002/jms.1614

John, C. M., & Bowen, D. (2016). Community software for challenging isotope analysis: First applications of ‘Easotope’ to clumped isotopes. *Rapid Communications in Mass Spectrometry*, *30*(21), 2285–2300. https://doi.org/10.1002/rcm.7720

Katz, A., Bonifacie, M., Hermoso, M., Cartigny, P., & Calmels, D. (2017). Laboratory-grown coccoliths exhibit no vital effect in clumped isotope (Δ47) composition on a range of geologically relevant temperatures. *Geochimica et Cosmochimica Acta*, *208*, 335–353. https://doi.org/10.1016/J.GCA.2017.02.025

Kim, S. T., & O’Neil, J. R. (1997). Equilibrium and nonequilibrium oxygen isotope effects in synthetic carbonates. *Geochimica et Cosmochimica Acta*, *61*(16), 3461–3475. https://doi.org/10.1016/S0016-7037(97)00169-5

Kele, S., Breitenbach, S. F. M., Capezzuoli, E., Nele Meckler, A., Ziegler, M., Millan, I. M., et al. (2015). Temperature dependence of oxygen- and clumped isotope fractionation in carbonates: a study of travertines and tufas in the 6-95°C temperature range. *Geochimica et Cosmochimica Acta*, *168*, 172–192. https://doi.org/10.1016/j.gca.2015.06.032

Kelson, J. R., Huntington, K. W., Schauer, A. J., Saenger, C., & Lechler, A. R. (2017). Toward a universal carbonate clumped isotope calibration: Diverse synthesis and preparatory methods suggest a single temperature relationship. *Geochimica et Cosmochimica Acta*, *197*. https://doi.org/10.1016/j.gca.2016.10.010

Kluge, T., and Affek, H.P. (2012) Quantifying kinetic fractionation in Bunker Cave speleothems using Δ_{47}, *Quaternary Science Reviews*, *49*, 82-94, https://dx.doi.org/10.1016/j.quascirev.2012.06.013

Kluge, T., John, C. M., Jourdan, A.-L., Davis, S., & Crawshaw, J. (2015). Laboratory calibration of the calcium carbonate clumped isotope thermometer in the 25–250°C temperature range. *Geochimica et Cosmochimica Acta*, *157*, 213–227. https://doi.org/10.1016/j.gca.2015.02.028

Kluge, T., & John, C. M. (2015). Effects of brine chemistry and polymorphism on clumped isotopes revealed by laboratory precipitation of mono- and multiphase calcium carbonates. *Geochimica et Cosmochimica Acta*, *160*, 155–168. https://doi.org/10.1016/J.GCA.2015.03.031

LeGrande, A. N., & Schmidt, G. A. (2006). Global gridded data set of the oxygen isotopic composition in seawater. *Geophysical Research Letters*, *33*(12). https://doi.org/10.1029/2006GL026011

Levitt, N.P., Eiler, J. M., Romanek, C. S., Beard, B. L., Xu, H., & Johnson C. M. (2018). Near Equilibrium 13C-18O Bonding During Inorganic Calcite Precipitation Under Chemo-Stat Conditions. *Geochemistry, Geophysics, Geosystems*, 313(14), 309–20.

Li, W., Ni, B., Jin, D., & Zhang, Q. (1988). Measurement of the absolute abundance of oxygen-17 in V-SMOW. *Chinese Science Bulletin*, *33*(19), 1610–1613. Retrieved from http://ci.nii.ac.jp/naid/80004607415/en/

Luz, B., & Barkan, E. (2010). Variations of 17O/16O and 18O/16O in meteoric waters. *Geochimica et Cosmochimica Acta*, *74*(22), 6276–6286. https://doi.org/10.1016/j.gca.2010.08.016

Matsuhisa, Y., Goldsmith, J. R., & Clayton, R. N. (1978). Mechanisms of hydrothermal crystallization of quartz at 250°C and 15 kbar. *Geochimica et Cosmochimica Acta*, *42*(2), 173–182. https://doi.org/10.1016/0016-7037(78)90130-8

McCrea, J. M. (1950). On the Isotopic Chemistry of Carbonates and a Paleotemperature Scale. *The Journal of Chemical Physics*, *18*(849), 849–857. https://doi.org/10.1063/1.1747785

Meckler, A. N., Ziegler, M., Millán, M. I., Breitenbach, S. F. M., & Bernasconi, S. M. (2014). Long-term performance of the Kiel carbonate device with a new correction scheme for clumped isotope measurements. *Rapid Communications in Mass Spectrometry*, *28*(15), 1705–1715. https://doi.org/10.1002/rcm.6949

Müller, I. A., Fernandez, A., Radke, J., van Dijk, J., Bowen, D., Schwieters, J., & Bernasconi, S. M. (2017). Carbonate clumped isotope analyses with the long-integration dual-inlet (LIDI) workflow: scratching at the lower sample weight boundaries. *Rapid Communications in Mass Spectrometry*, *31*(12). https://doi.org/10.1002/rcm.7878

Murray, S. T., Arienzo, M. M., & Swart, P. K. (2016). Determining the Δ47 acid fractionation in dolomites. *Geochimica et Cosmochimica Acta*, *174*, 42–53. https://doi.org/10.1016/J.GCA.2015.10.029

Passey, B. H., Levin, N. E., Cerling, T. E., Brown, F. H., & Eiler, J. M. (2010). High-temperature environments of human evolution in East Africa based on bond ordering in paleosol carbonates. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(25), 11245–9. https://doi.org/10.1073/pnas.1001824107

Passey, B. H., & Henkes, G. A. (2012). Carbonate clumped isotope bond reordering and geospeedometry. *Earth and Planetary Science Letters*, *351–352,* 223–236. https://doi.org/10.1016/j.epsl.2012.07.021

Peral, M., Daëron, M., Blamart, D., Bassinot, F., Dewilde, F., Smialkowski, N., et al. (2018). Updated calibration of the clumped isotope thermometer in planktonic and benthic foraminifera. *Geochimica et Cosmochimica Acta*, *239*, 1-16. https://doi.org/10.1016/J.GCA.2018.07.016

Petersen, S. V., & Schrag, D. P. (2014). Clumped isotope measurements of small samples using a high-efficiency dual-reservoir technique. *Rapid Communications in Mass Spectrometry, 28,* 2371-2381. http://doi.org/10.1002/rcm.7022

Petersen, S. V., Winkelstern, I. Z., Lohmann, K. C., & Meyer, K. W. (2016). The effects of Porapak™ trap temperature on d18O, d13C, and D47 values in preparing samples for clumped isotope analysis. *Rapid Communications in Mass Spectrometry, 30*, 199-208. https://doi.org/10.1002/rcm.7438

Petrizzo, D. A., Young, E. D., & Runnegar, B. N. (2014). Implications of high-precision measurements of 13C–18O bond ordering in CO2 for thermometry in modern bivalved mollusc shells. *Geochimica et Cosmochimica Acta*, *142*, 400–410. https://doi.org/10.1016/j.gca.2014.07.017

Petrizzo, D. A., & Young, E. D. (2014). High-precision determination of 13C–18O bonds in CO2 using multicollector peak hopping. *Rapid Communications in Mass Spectrometry*, *28*(11), 1185–1193. https://doi.org/10.1002/rcm.6888

Rosenheim, B. E., Tang, J., & Fernandez, A. (2013). Measurement of multiply substituted isotopologues ('clumped isotopes’) of CO2 using a 5 kV compact isotope r
spectrometer: Performance, reference frame, and carbonate paleothermometry. *Rapid Communications in Mass Spectrometry*, *27*(16), 1847–1857. https://doi.org/10.1002/rcm.6634

Saenger, C., Affek, H. P., Felis, T., Thiagarajan, N., Lough, J. M., & Holcomb, M. (2012). Carbonate clumped isotope variability in shallow water corals: Temperature dependence and growth-related vital effects. *Geochimica et Cosmochimica Acta*, *99,* 224–242. https://doi.org/10.1016/j.gca.2012.09.035

Saenger, C., Gabitov, R. I., Farmer, J., Watkins, J. M., & Stone, R. (2017). Linear correlations in bamboo coral d13C and d18O sampled by SIMS and micromill: Evaluating paleoceanographic potential and biomineralization mechanisms using d11B and D47 composition. *Chemical Geology, 454,* 1-14. https://doi.org/10.1016/j.chemgeo.2017.02.014

Sample, J.C., Torres, M. E., Fisher, A., Hong, W-L., Destrigneville, C., Defliese, W. F., & Tripati, A. E. (2017). Geochemical constraints on the temperature and timing of carbonate formation and lithification in the Nankai Trough, NanTroSEIZE transect. *Geochimica et Cosmochimica Acta, 198,* 92–114.

Santrock, J., Studley, S. A., & Hayes, J. M. (1985). Isotopic analyses based on the mass spectrum of carbon dioxide. *Analytical Chemistry*, *57,* 1444–1448. https://doi.org/10.1021/ac00284a060

Schauble, E. A., Ghosh, G., & Eiler, J. M. (2006). Preferential formation of 13C-18O bonds in carbonate minerals, estimated using first-principles lattice dynamics. *Geochimica et Cosmochimica Acta*, *70*, 2510-2529. https://doi.org/10.1016/j.gca.2006.02.011

Schauer, A. J., Kelson, J., Saenger, C., & Huntington, K. W. (2016). Choice of 17O correction affects clumped isotope (Δ47) values of CO2 measured with mass spectrometry. *Rapid Communications in Mass Spectrometry*, *30*(24), 2607–2616. https://doi.org/10.1002/rcm.7743

Schmid, T. W., & Bernasconi, S. M. (2010). An automated method for “clumped-isotope” measurements on small carbonate samples. *Rapid Communications in Mass Spectrometry*, *23*, 1955–1963. https://doi.org/10.1002/rcm.4598

Spooner, P. T., Guo, W., Robinson, L. F., Thiagarajan, N., Hendry, K. R., Rosenheim, B. E., & Leng, M. J. (2016). Clumped isotope composition of cold-water corals: A role for vital effects? *Geochimica et Cosmochimica Acta*, *179*, 123–141. https://doi.org/10.1016/J.GCA.2016.01.023

Staudigel, P. T., & Swart, P. K. (2016). Isotopic behavior during the aragonite-calcite transition: Implications for sample preparation and proxy interpretation. *Chemical Geology*, *442*, 130–138. https://doi.org/10.1016/J.CHEMGEO.2016.09.013

Stolper, D. A., & Eiler, J. M. (2015). The kinetics of solid-state isotope-exchange reactions for clumped isotopes: A study of inorganic calcites and apatites from natural and experimental samples. *American Journal of Science*, *315*(5), 363–411. https://doi.org/10.2475/05.2015.01

Tang, J., Dietzel, M., Fernandez, A., Tripati, A. K., & Rosenheim, B. E. (2014). Evaluation of kinetic effects on clumped isotope fractionation (Δ47) during inorganic calcite precipitation. *Geochimica et Cosmochimica Acta*, *134*, 120–136. https://doi.org/10.1016/j.gca.2014.03.005

Thiagarajan, N., Adkins, J., & Eiler, J. (2011). Carbonate clumped isotope thermometry of deep-sea corals and implications for vital effects. *Geochimica et Cosmochimica Acta*, *75*(16), 4416–4425. https://doi.org/10.1016/j.gca.2011.05.004

Tripati, A. K., Eagle, R. A., Thiagarajan, N., Gagnon, A. C., Bauch, H., Halloran, P. R., & Eiler, J. M. (2010). 13C–18O isotope signatures and ‘clumped isotope’ thermometry in foraminifera and coccoliths. *Geochimica et Cosmochimica Acta*, *74*(20), 5697–5717. https://doi.org/10.1016/j.gca.2010.07.006

Tripati, A. K., Hill, P. S., Eagle, R. A., Mosenfelder, J. L., Tang, J., Schauble, E. A., et al. (2015). Beyond temperature: Clumped isotope signatures in dissolved inorganic carbon species and the influence of solution chemistry on carbonate mineral composition. *Geochimica et Cosmochimica Acta*, *166*, 344-371. https://dx.doi.orng/10.1016/j.gca.2015.06.021

van Dijk J., Fernandez A., Storck J. C, White T., Lever M., Müller I.A., et al. (2019) Experimental calibration of clumped isotopes in siderite between 8.5 and 62°C and its application as paleo-thermometer in paleosols. *Geochimica et Cosmochimica Acta*, *254*, 1-20. https://doi.org/10.1016/j.gca.2019.03.018

Ext. 1 sd[8]
0.022
0.022
0.020
0.020
0.015
0.025
0.017
0.025
0.016
0.017
0.016
0.022
0.020
0.014
0.040

Venturelli, R.A. and Rosenheim B.E. (2019) Compositional and beam-size-dependent effects on pressure baseline in clumped isotope mass spectrometry. *Rapid Communications in Mass Spectrometry, 33,* 140-148, https://doi.org/10.1002/rcm.8303

Wacker, U., Fiebig, J., Tödter, J., Schöne, B. R., Bahr, A., Friedrich, O., et al. (2014). Empirical calibration of the clumped isotope paleothermometer using calcites of various origins. *Geochimica et Cosmochimica Acta, 141*, 127–144. https://doi.org/10.1016/j.gca.2014.06.004

Wacker, U., Fiebig, J., & Schoene, B. R. (2013). Clumped isotope analysis of carbonates: comparison of two different acid digestion techniques. *Rapid Communications in Mass Spectrometry*, *27*(14), 1631–42. https://doi.org/10.1002/rcm.6609

Wang, Z., Schauble, E. A., & Eiler, J. M. (2004). Equilibrium thermodynamics of multiply substituted isotopologues of molecular gases. *Geochimica et Cosmochimica Acta*, *68*(23), 4779–4797. https://doi.org/10.1016/j.gca.2004.05.039

Webb, M.A. & Miller, T.F., III (2014). Position-Specific and Clumped Stable Isotope Studies: Comparison of the Urey and Path-Integral Approaches for Carbon Dioxide, Nitrous Oxide, Methane, and Propane. *The Journal of Physical Chemistry A*, *118(2)*, 467–474.

Winkelstern, I. Z., Kaczmarek, S. E., Lohmann, K. C., & Humphrey, J. D. (2016). Calibration of dolomite clumped isotope thermometry. *Chemical Geology, 443*, 32–38. https://doi.org/10.1016/J.CHEMGEO.2016.09.021

Zaarur, S., Olack, G., & Affek, H. P. (2011). Paleo-environmental implication of clumped isotopes in land snail shells. *Geochimica et Cosmochimica Acta*, *75*(22), 6859–6869. https://doi.org/10.1016/J.GCA.2011.08.044

Zaarur, S., Affek, H. P., & Brandon, M. T. (2013). A revised calibration of the clumped isotope thermometer. *Earth and Planetary Science Letters, 382,* 47–57. https://doi.org/10.1016/j.epsl.2013.07.026

University (Lab PI)	Sample type[1]	Mineralogy[2]	Reaction vessel type[3]	Acid temp.	Cleaning method[4]	Mass spectrometer	Prep. Device[5]	Raw data format[6]	PBL[7]	Standards for ARF calculation	Correction Method
University of Michigan	SYN	CA	CAB	75°C	Static PPQ	Thermo MAT253	MOL	raw voltages	N	HG and 25°C EG	Variable-length moving window
University of Michigan	SYN nat	D	CAB	75°C	Static PPQ	Thermo MAT253	MOL	raw voltages	N	HG and 25°C EG	Variable-length moving window
Imperial College London	SYN	CAV	IND	70°C and 90°C	Static PPQ	Thermo MAT253	MOL	smδ Easotope	N	HG, 20°C, 50°C, 80°C EGs, ETH-1, -3 -,4, Carrara	Long fixed windows, HG only for SlopeEGL, gases + carb stds for ETF
Imperial College London	SYN nat	M	IND	90°C	Static PPQ	Thermo MAT253	MOL	smδ Easotope	N	HG, 20°C, 50°C, 80°C EGs, ETH-1, -3 -,4, Carrara	Long fixed windows, HG only for SlopeEGL
University of Washington	SYN	CA	CAB (most), IND (MC) for 25°C reactions	25°C and 90°C	He-carrier PPQ	Thermo MAT253	AOL	smδ	Y	HG, 60°C and 4°C EGs	Long fixed windows
Tulane	SYN	C	IND	100°C	Static PPQ	Elementar Isoprime 100	MOL	raw voltages	Some	HG, 51°C, 28°C, and 4°C EGs	Modeled daily HG slope, long fixed ETF window
UCLA (Tripati)			CAB	90°C	GC	Thermo MAT253	APD	smδ Easotope	N	HG and 25°C EG	Long fixed window
Tulane	SYN	S	IND	100°C	Static PPQ	Elementar Isoprime 100	MOL	raw voltages	Some	HG, 51°C, 28°C, and 4°C EGs	Modeled daily HG slope, long fixed ETF window
Johns Hopkins University	SYN	C	CAB	90°C	GC	Thermo MAT253	APD	raw voltages	N	HG, and 25°C or 30°C EG	Daily modeled HG slope and ETF
LSCE	BIO	C	CAB	90°C	He-carrier PPQ	Elementar Isoprime 100	MOL	smδ	N	ETH-1, -2, and -3	Single-step solve, Long fixed windows.
Johns Hopkins University	BIO	CA	CAB	90°C	GC	Thermo MAT253	APD	raw voltages	N	HG and 25°C or 30°C EG	Daily modeled HG slope and ETF
Goethe University	BIO nat	C	CAB	90°C	GC	Thermo MAT253	MOL	smδ Easotope	N	HG and 25°C EG	Variable-length moving window for SlopeEGL, long fixed window for ETF
UCLA (Young)	BIO	CA	IND (MC)	25°C	GC	Thermo MAT253 w/ peak hopping	MOL	raw voltages	Y	HG, 25°C and 2°C EGs	Long fixed windows
IPGP	BIO	C	CAB	90°C	Static PPQ	Thermo MAT253	MOL	smδ	N	HG and 25°C EG	Long fixed windows
University of Cambridge	BIO nat	C	IND	70°C	Static PPQ	Thermo MAT253	AKD	smδ	Y	ETH-1, -2, -3, and -4	ETF only, long fixed windows

Study	*Defliese et al., 2015*	*Winkelstern et al., 2016*	*Kluge et al., 2015*	*Garcia de Real et al., 2016*	*Kelson et al., 2017*	*Tang et al., 2014*	*Fernandez et al., 2014*	*Passey and Henkes, 2012*	*Peral et al., 2018*	*Henkes et al., 2013*	*Wacker et al., 2014*	*Petrizzo et al., 2014*	*Katz et al., 2017*	*Breitenbach et al., 2018*
#	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Table 1. Summary of sample preparation information and data processing information for published calibration datasets reprocessed as part of this study. ^{1}Sample type: SYN=synthetic, nat=natural inorganic, BIO=biogenic. ^{2}Mineralogy: C=Calcite, A=aragonite, V=vaterite, M=magnesite, S=siderite, D=dolomite. ^{3}CAB = common acid bath, IND = individual reaction vessels, MC=McCrea-type. ^{4}Cleaning Method: Static PPQ = no carrier gas, hand packed column, He-carrier PPQ = He carrier gas, hand-packed column, GC=gas chromatograph with He carrier gas. ^{5}Preparation Device: MOL = manual off-line, APD= automated "Passey" device, AOL = automated off-line, AKD = automated Kiel-based device. ^{6}Raw data format: smδ = small delta values, smδ Easotope = small delta values from Easotope. ^{7}PBL = Pressure Baseline: Yes/No/Some. HG = heated gas, EG= equilibrated gas. ^{8}Ext. 1sd = external 1 standard deviation on a single replicate.

Acid temp.	**Guo et al., 2009**	**Defliese et al., 2015**	**Reprocessed data** $\Delta_{47\text{-RFAC (SG,WD)}}$ **C/A/D (*n*=151)**	**Reprocessed data** $\Delta_{47\text{-RFAC (Br,P)}}$ **C/A/D (*n*=151)****	**Reprocessed data** $\Delta_{47\text{-RFAC (Br,P)}}$ **calcite only (*n*=87)**	**Reprocessed data** $\Delta_{47\text{-RFAC (Br,P)}}$ **aragonite only (*n*=48)**	**Reprocessed data** $\Delta_{47\text{-RFAC (Br,P)}}$ **dolomite only (*n*=16)**
100°C	0.077‰	0.091‰	0.105 ± 0.006‰	0.098 ± 0.006‰	0.094 ± 0.008‰	0.104 ± 0.013‰	0.098 ± 0.011‰
90°C	0.069‰	0.082‰	0.095 ± 0.006‰	0.088 ± 0.006‰	0.085 ± 0.008‰	0.094 ± 0.011‰	0.089 ± 0.010‰
75°C	0.057‰	0.067‰	0.078 ± 0.005‰	0.072 ± 0.005‰	0.069 ± 0.008‰	0.077 ± 0.009‰	0.072 ± 0.008‰
70°C	0.052‰	0.062‰	0.071 ± 0.004‰	0.066 ± 0.004‰	0.064 ± 0.008‰	0.071 ± 0.008‰	0.067 ± 0.007‰

Table 2. Summary of $\Delta^*_{25\text{-}90}$ values discussed in Section 3.3, including published values from Guo et al. (2009) (Eqn. 23) and Defliese et al. (2015) (Eqn. 2) and reprocessed values for relevant replicates of all available mineralogies (C/A/D) or a single mineralogy (calcite, aragonite, or dolomite). ** indicates recommended $\Delta^*_{25\text{-}90}$ values to use in future studies. See **Table S3** for $\Delta^*_{25\text{-}90}$ values calculated with different parameter combinations.

Data	**Slope**	**Slope SE**	**Intercept**	**Intercept SE**	r^2	**p-value**
Only synthetic, $\Delta_{47\text{-RFAC (SG,WD,newAFF)}}$	0.0370	1.7E-06	0.281	1.7E-05	0.92	<<0.0001
Only synthetic, $\Delta_{47\text{-RFAC (SG,P,newAFF)}}$	0.0368	1.7E-06	0.280	1.7E-05	0.92	<<0.0001
Only synthetic, $\Delta_{47\text{-RFAC (Br,WD,newAFF)}}$	0.0387	1.7E-06	0.257	1.7E-05	0.94	<<0.0001

Only synthetic, $\Delta_{47\text{-RFAC (Br,P,newAFF)}}$	**0.0383**	**1.7E-06**	**0.258**	**1.7E-05**	**0.94**	**<<0.0001**

Table 3. Summary of Monto Carlo regression parameters (mean values and errors) for regressions through all synthetic replicates, corrected using Santrock/Gonfiantini (SG) or IUPAC (Br) parameters, Wang/Dennis (WD) or Petersen (P) $\Delta_{47\text{-TE}}$ values, and new reprocessed $\Delta^*_{25\text{-X}}$ values (newAFF) (see **Figure 1** for abbreviations). Errors on slopes and intercepts are 1 SE taken from Monto Carlo sampling (see Section 2.4). See **Tables S4 and S5** for regressions through other subsets of data (synthetic+biogenic, formation temperature <100°C, individual laboratories).

Figure 1. Diagram describing inputs, outputs, and processing steps in the data reprocessing workflow used in this study. Abbreviations and subscripts used throughout the manuscript are assembled here for reference.

Figure 2. Change in fully-corrected Δ_{47} ($\Delta\Delta_{47\text{-RF}}$) for all sample replicates as the result of updating from Wang/Dennis $\Delta_{47\text{-TE}}$ values to new $\Delta_{47\text{-TE}}$ values (this study) (WD minus P), separated by study. Colors correspond to study, with first author name listed in corresponding color (more details given in **Table 1**). Data shown here was corrected using IUPAC parameters (Br), but results are nearly identical for data corrected with Santrock/Gonfiantini parameters ($\Delta\Delta_{47\text{-RF}}$ WD-P, SG). Studies showing no change use only carbonate standards or carbonate standards plus heated gases to calculate ARF parameters. No acid digestion fractionation correction included to isolate effects of changing $\Delta_{47\text{-TE}}$ values.

Figure 3. Change in fully-corrected Δ_{47} ($\Delta\Delta_{47\text{-RF}}$) for all sample replicates as the result of updating from Santrock/Gonfiantini to IUPAC parameters (SG-Br), separated by study **(a)** or plotted against IUPAC $\delta^{13}C$ **(b)** and $\delta^{18}O$ **(c)**. Colors in all panels correspond to study, with first author name listed in corresponding color in (a) and more details given in Table 1. Data shown here was corrected using the updated $\Delta_{47\text{-TE}}$ values (P, this study), but results are nearly identical for data corrected with Wang/Dennis $\Delta_{47\text{-TE}}$ values (WD). Plots (b) and (c) are expanded in **Figures S2 and S3**, separated by individual study.

Figure 4. Composite calibration line (Monte Carlo fit) through sample mean $\Delta_{47\text{-RFAC}}$ values **(a/c)** and offset from corresponding composite calibration ($\Delta\Delta_{47\text{-RFAC}}$) **(b/d)**, selecting synthetic carbonates only, processed using either Santrock/Gonfiantini **(a/b)** or IUPAC **(c/d)** parameters, new $\Delta_{47\text{-TE}}$ values (Petersen, this study) and new $\Delta^*_{25\text{-X}}$ values (this study). Error bars represent 1SE external error on the sample mean Δ_{47} (vertical) or reported error in formation temperature (horizontal). Colors correspond to study, with first author name listed in corresponding color

(more details given in **Table 1**). Theoretical predictions of clumping in solid calcite by Schauble et al. (2006) and calcite and aragonite by Hill et al. (2014) combined with experimentally or theoretically determined fractionation between solid calcite or aragonite, respectively, and CO_2 from Guo et al. (2009) updated to the absolute reference frame ($\Delta^*_{calcite}$ = +0.232 becomes +0.268‰, $\Delta^*_{aragonite}$ = +0.229‰ becomes +0.264‰ using Table 4 of Dennis et al., 2011) shown for comparison, although these inherently include an estimation of $\Delta^*_{25\text{-}X}$ that has not been updated to IUPAC parameters. This would only differ by 0.004‰ if updated using $\Delta^*_{dolomite90}$ from Bonifacie et al. (2017) and our +0.088‰ $\Delta^*_{25\text{-}90}$ value (**Table 2**). Linear fit through all data calculated using Monte Carlo sampling approach, with statistics and fit parameters listed in **Table 2** and shown in lower right of panels a and c. Total residuals are calculated as the sum of the square of the absolute value of the offset between observed and predicted from the synthetic composite calibration.

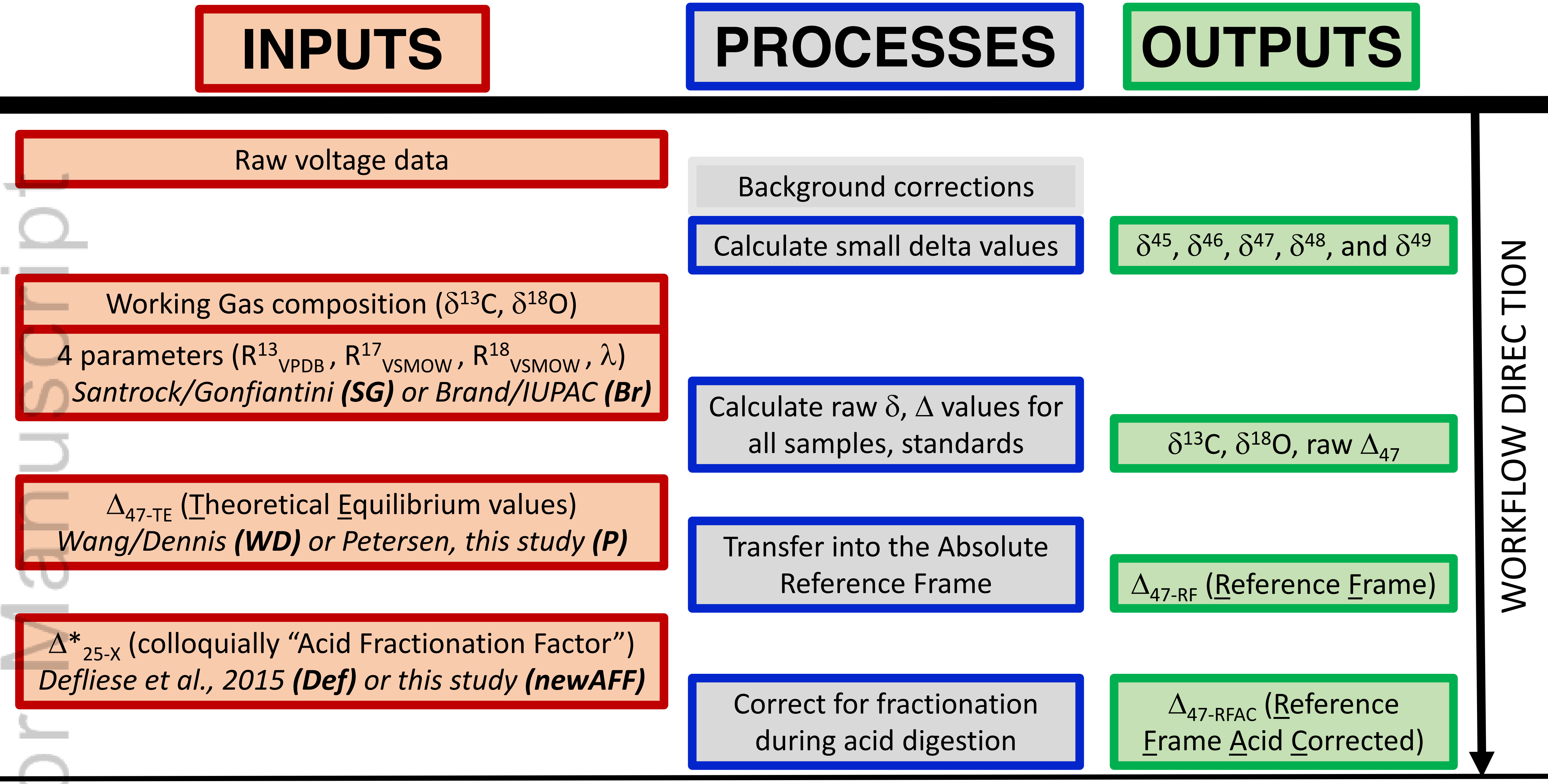
Figure 5. Offset between non-calcite sample mean $\Delta_{47\text{-}RFAC}$ values and an ordinary least squares regression through all calcite samples (synthetic, biogenic, and natural inorganic), processed using **(a)** SG or **(b)** IUPAC parameters and new $\Delta_{47\text{-}TE}$ and $\Delta^*_{25\text{-}X}$ values (Petersen, this study). The same set of $\Delta^*_{25\text{-}X}$ values have been applied to all mineralogies. Error bars correspond to 1 SE external error on sample mean Δ_{47} (vertical) or reported uncertainty in formation temperature (horizontal). Colors correspond to mineralogy (calcite, aragonite, dolomite, vaterite, magnesite, siderite, abbreviated by their first letter), and symbols correspond to sample type (synthetic, natural inorganic, biogenic). Colored-coded lines and shading depict a linear fit and 95% confidence interval for each mineralogy.

Figure 6. 95% confidence ellipses for the calibration slope and intercept value estimated at 25°C of each reprocessed study using IUPAC parameters and Petersen $\Delta_{47\text{-}TE}$ values (left) vs. SG parameters and WD $\Delta_{47\text{-}TE}$ values (right). Updated $\Delta^*_{25\text{-}X}$ values are used in both cases. First author name is plotted in the centroid of each ellipse. Using $\Delta_{47\text{-}RFAC(Br,P,newAFF)}$ values, at most 9 of 14 studies can be in agreement at once.

Figure 7. Offset between sample mean $\Delta_{47\text{-}RFAC}$ values and the synthetic composite calibration (**Equation 1**), processed using IUPAC parameters, new $\Delta_{47\text{-}TE}$ values (Petersen) and new $\Delta^*_{25\text{-}X}$ values (this study), separated by study, with 1 SE external standard error bars. Colors correspond to study, with first author name listed in corresponding color (more details given in **Table 1**). A few natural inorganic points from Winkelstern and Garcia are included in the synthetic section with the majority of the data from those studies. Data from Kele et al. (2015) was updated using IUPAC parameters by Bernasconi et al. (2018) and three additional samples were included ("Kele+"). Bonifacie et al. (2017) data uses older SG parameters, but is not expected to change much based on the limited change in data from Katz et al. (2017) and the composition of standards used at IPGP. Both Kele+ and Bonifacie data have been updated to use new $\Delta^*_{25\text{-}X}$

values. In grey panel, the full range in scatter around the synthetic composite calibration combining all 14 reprocessed studies is displayed (Range Br/P) and compared to the larger range for the same data processed using SG parameters, WD $\Delta_{47\text{-TE}}$ values and new $\Delta^*_{25\text{-X}}$ values (Range SG/WD).

Figure 8. Formation temperature vs. error on reconstructed temperature (1SE) on a log scale showing the relative contributions of different sources of error to total uncertainty on a reconstructed clumped isotope temperature. Assuming a long-term reproducibility of 0.020‰ (1sd) on Δ_{47} (which is achieved by most labs - see **Table 1**), red, green, and blue lines show the purely instrumental or analytical error (ignoring all calibration uncertainties) 1SE error for N=1, 4, or 10 replicates at different formation temperatures. Black lines show the error introduced by the calibration (from uncertainty in slope and intercept values, computed in the same way as in **Figure 6**) for four representative calibrations spanning different temperature ranges and containing different numbers of replicate analyses. For well-constrained calibrations (e.g. N>30 replicates), error from uncertainty in the slope and intercept of the calibration is much less than the analytical error, even with >10 replicates of the unknown.



INPUTS **PROCESSES** **OUTPUTS**

Raw voltage data

Background corrections

Calculate small delta values

δ^{45}, δ^{46}, δ^{47}, δ^{48}, and δ^{49}

Working Gas composition ($\delta^{13}C$, $\delta^{18}O$)

4 parameters (R^{13}_{VPDB}, R^{17}_{VSMOW}, R^{18}_{VSMOW}, λ)
*Santrock/Gonfiantini **(SG)** or Brand/IUPAC **(Br)***

Calculate raw δ, Δ values for all samples, standards

$\delta^{13}C$, $\delta^{18}O$, raw Δ_{47}

$\Delta_{47\text{-}TE}$ (Theoretical Equilibrium values)
*Wang/Dennis **(WD)** or Petersen, this study **(P)***

Transfer into the Absolute Reference Frame

$\Delta_{47\text{-}RF}$ (Reference Frame)

$\Delta^{*}_{25\text{-}X}$ (colloquially "Acid Fractionation Factor")
*Defliese et al., 2015 **(Def)** or this study **(newAFF)***

Correct for fractionation during acid digestion

$\Delta_{47\text{-}RFAC}$ (Reference Frame Acid Corrected)

RESULT: $\Delta_{47\text{-}RFAC}$ (parameter set, theoretical equilibrium value set, acid fractionation factor value)

e.g. $\Delta_{47\text{-}RFAC}$ (Br, P, newAFF) = Δ_{47} values calculated with IUPAC parameters, new $\Delta_{47\text{-}TE}$ and $\Delta^{*}_{25\text{-}X}$ values, corrected into the absolute reference frame and presented relative to a 25°C acid reaction

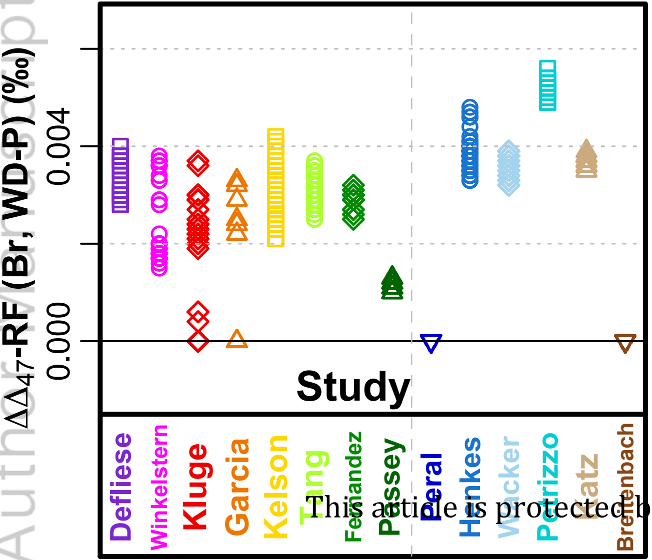
Other abbreviations:

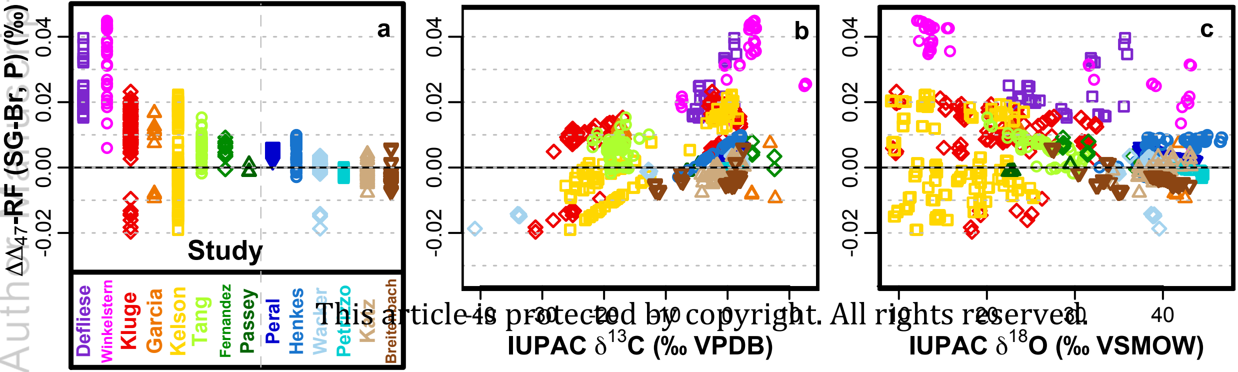
$\Delta\Delta_{47\text{-}RFAC}$ = offset in Δ_{47} values from a stated calibration equation (e.g. calcite-only or synthetics-only)

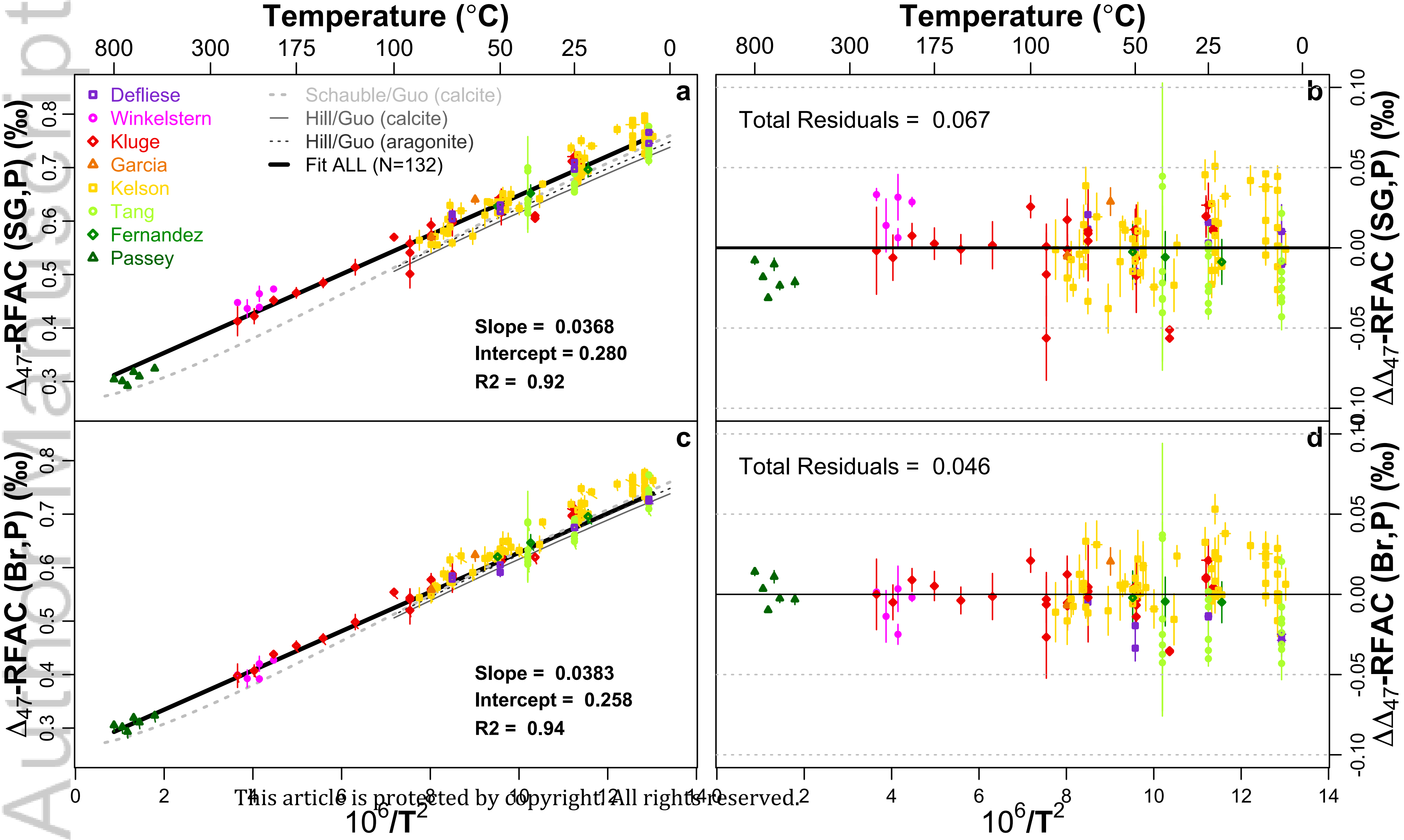
$\Delta\Delta_{47\text{-}RFAC}$ (X-Y) = difference in Δ_{47} values calculated using X vs. Y parameter set or $\Delta_{47\text{-}TE}$ value set (eg. SG-Br or WD-P)

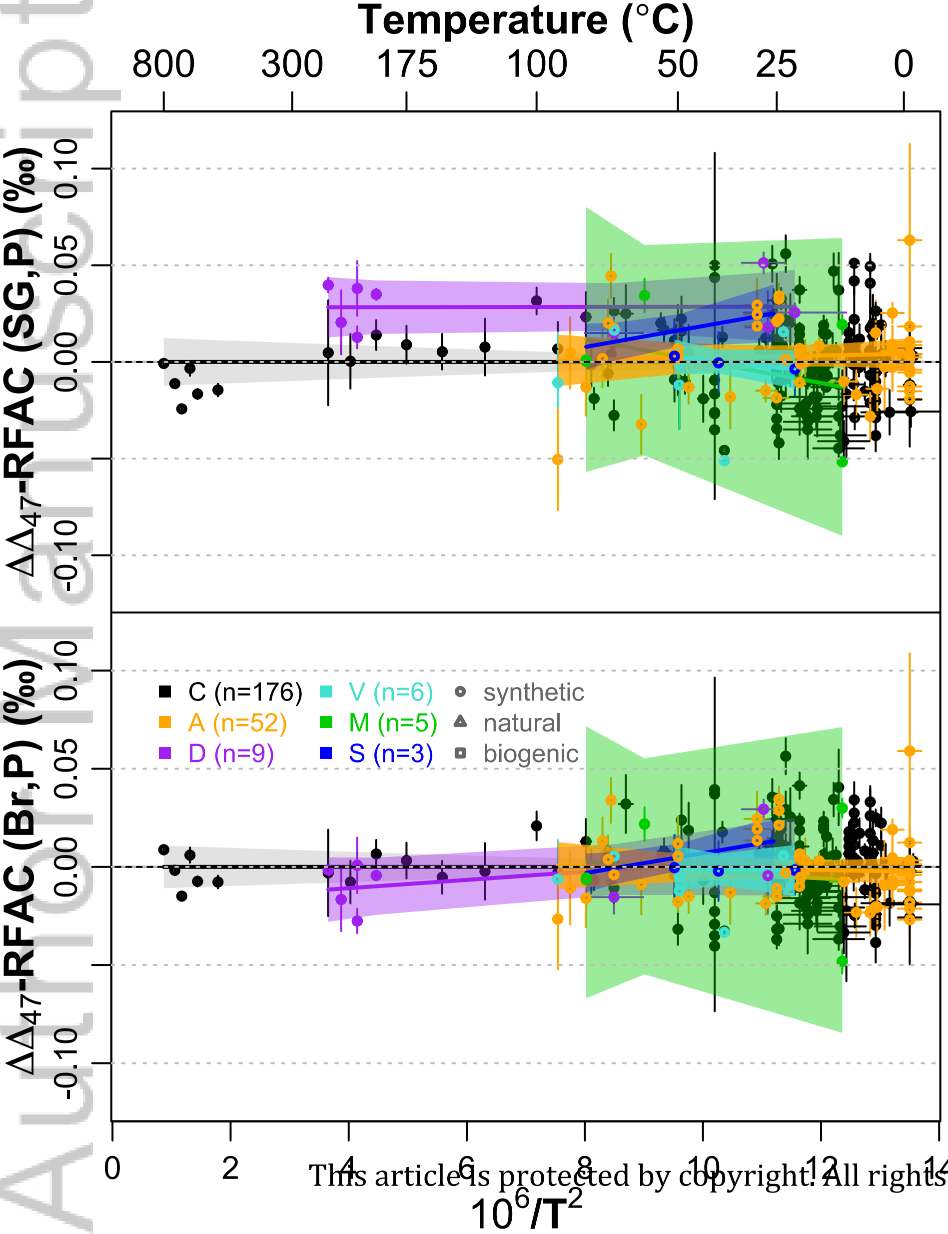
$\Delta^{*}_{mineralogyX}$ = Δ_{63} - Δ_{47}, or difference between clumping in solid carbonate and CO_2 produced by reaction at acid temperature X for a given mineralogy, also known as full or absolute acid fractionation factor

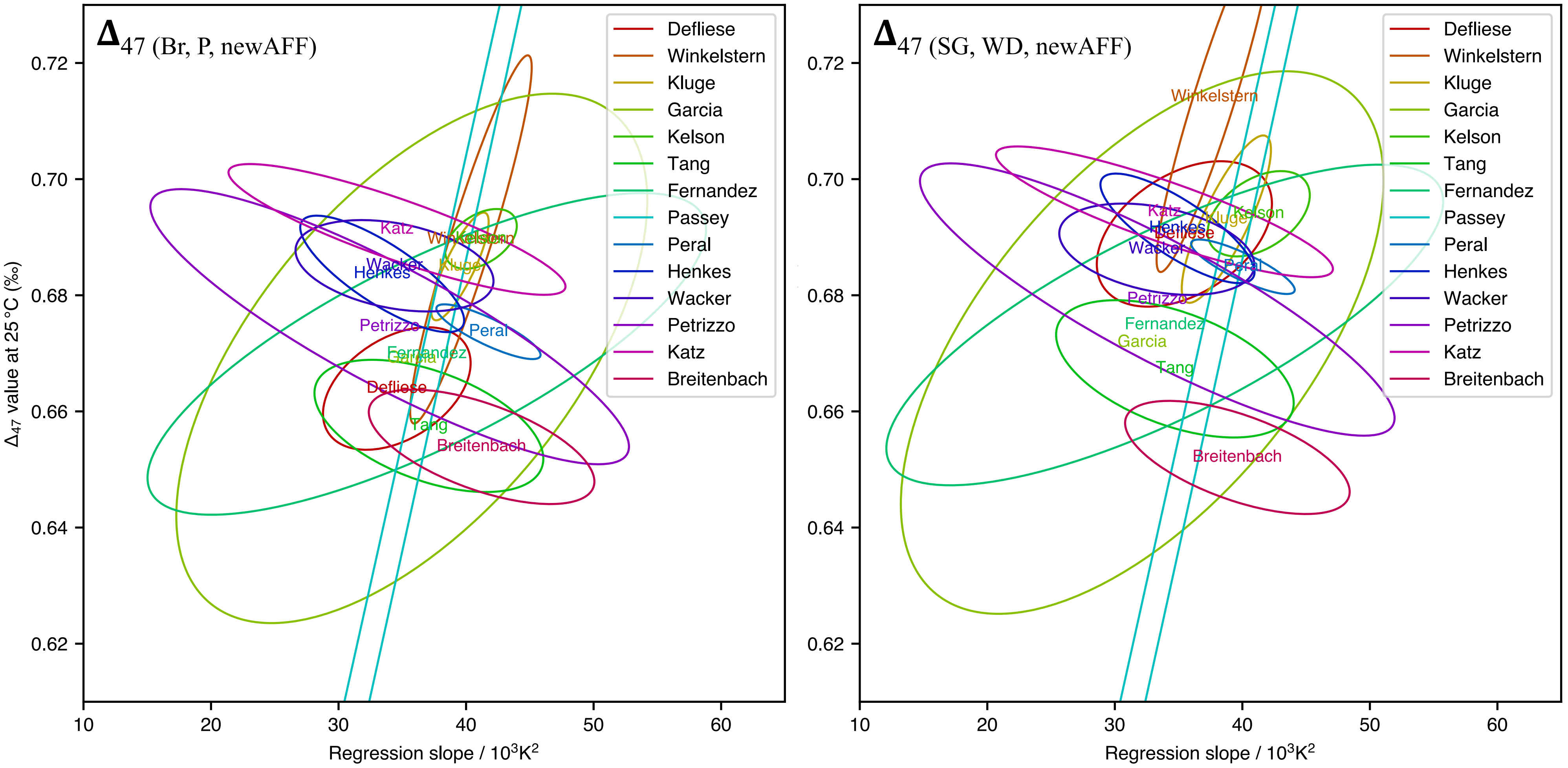
$\Delta^{*}_{25\text{-}X}$ = difference in clumping between CO_2 produced by reaction at acid temperature X, relative to a 25°C acid reaction, colloquially called the acid fractionation factor (AFF)


ΔΔ$_{47}$-RF (Br, WD-P) (‰)
0.000
0.004
Study
Defliese
Winkelstern
Kluge
Garcia
Kelson
Tang
Fernandez
Passey
Peral
Henkes
Wacker
Petrizzo
Katz
Breitenbach

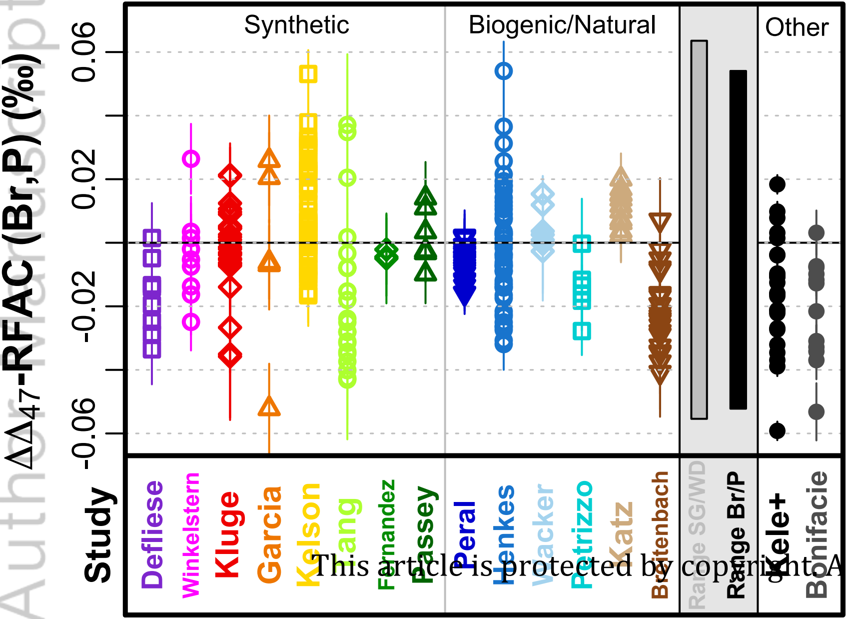

ΔΔ$_{47}$-RF (SG-Br, P) (‰)
0.04
0.02
0.00
-0.02
a
Study
Defliese
Winkelstern
Kluge
Garcia
Kelson
Tang
Fernandez
Passey
Peral
Henkes
Wacker
Petrizzo
Katz
Breitenbach
b
-40
-30
-20
-10
0
10
IUPAC δ^{13}C (‰ VPDB)
c
10
20
30
40
IUPAC δ^{18}O (‰ VSMOW)

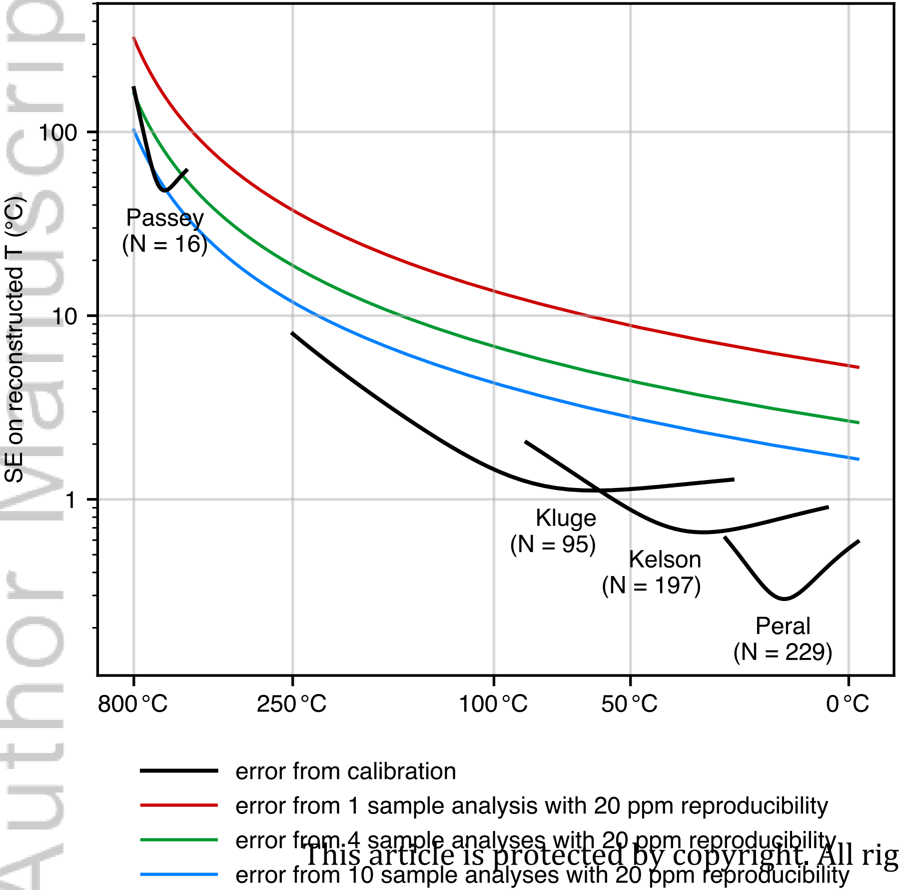

Temperature (°C)
800
300
175
100
50
25
0
a
Defliese
Winkelstern
Kluge
Garcia
Kelson
Tang
Fernandez
Passey
Schauble/Guo (calcite)
Hill/Guo (calcite)
Hill/Guo (aragonite)
Fit ALL (N=132)
Slope = 0.0368
Intercept = 0.280
R2 = 0.92
Δ_{47}-RFAC (SG,P) (‰)
0.3
0.4
0.5
0.6
0.7
0.8
b
Total Residuals = 0.067
$\Delta\Delta_{47}$-RFAC (SG,P) (‰)
-0.10
-0.05
0.00
0.05
0.10
c
Slope = 0.0383
Intercept = 0.258
R2 = 0.94
Δ_{47}-RFAC (Br,P) (‰)
d
Total Residuals = 0.046
$\Delta\Delta_{47}$-RFAC (Br,P) (‰)
0
2
4
6
8
10
12
14
$10^6/T^2$


Temperature (°C)
800
300
175
100
50
25
0
ΔΔ47-RFAC (SG,P) (‰)
0.10
0.05
0.00
-0.10
ΔΔ47-RFAC (Br,P) (‰)
0.10
0.05
0.00
-0.10
■ C (n=176)
■ A (n=52)
■ D (n=9)
■ V (n=6)
■ M (n=5)
■ S (n=3)
○ synthetic
△ natural
□ biogenic
0
2
4
6
8
10
12
1
10^6/T^2



Δ47 (Br, P, newAFF)
Δ47 (SG, WD, newAFF)
Δ47 value at 25 °C (‰)
0.72
0.70
0.68
0.66
0.64
0.62
10
20
30
40
50
60
Regression slope / 10³K²
Defliese
Winkelstern
Kluge
Garcia
Kelson
Tang
Fernandez
Passey
Peral
Henkes
Wacker
Petrizzo
Katz
Breitenbach


$\Delta\Delta_{47}$-RFAC (Br,P) (‰)
0.06
0.02
-0.02
-0.06
Synthetic
Biogenic/Natural
Other
Study
Defliese
Winkelstern
Kluge
Garcia
Kelson
Fang
Fernandez
Passey
Peral
Henkes
Wacker
Petrizzo
Katz
Breitenbach
Range SG/WD
Range Br/P
Kele+
Bonifacie


SE on reconstructed T (°C)
100
10
1
800 °C
250 °C
100 °C
50 °C
0 °C
Passey
(N = 16)
Kluge
(N = 95)
Kelson
(N = 197)
Peral
(N = 229)
error from calibration
error from 1 sample analysis with 20 ppm reproducibility
error from 4 sample analyses with 20 ppm reproducibility
error from 10 sample analyses with 20 ppm reproducibility