

Supplements to: *Nanometer-Scale Chemistry of a Calcite
Biomineralization Template: Implications for Skeletal
Composition and Nucleation*

O Branson E A Bonnin D E Perea H J Spero Z Zhu
M Winters B Hönlisch J S Fehrenbacher A D Russell A C Gagnon

DOI:10.1073/pnas.1522864113

Contents

1	Sample Preparation	3
1.1	Foraminifera Culturing	3
1.2	Sample Mounting and Polishing	3
1.3	FIB Preparation of APT Specimens	3
2	ToF-SIMS	6
2.1	Data Collection	6
2.2	Data Calibration	6
2.3	Data Processing	6
2.4	POS Identification	7
2.5	POS Measurement	9
3	Atom Probe Tomography (APT)	11
3.1	Specimen Optimization	11
3.2	Data Collection	11
3.3	Data Processing	12
3.4	Secondary Organic Regions	20
4	Scale-Up Model	21
4.1	Context	21
4.2	Hypothesis	21
4.3	Parameterization	22
4.4	Assumptions	24
4.5	Interpreting Model Output	24
4.6	Extension to Bulk Geochemistry	26
4.7	Convolution Calculation	27

5	Data Processing Code	29
5.1	Data	29
5.2	Python Environment	29
5.3	Compositional Analysis	29
5.4	Data Quality Checks	36
5.5	Proxygram Processing	37
5.6	APT Helper Functions	42

1 Sample Preparation

1.1 Foraminifera Culturing

Foraminiferal culture techniques have been described in detail elsewhere (for example, [Lea and Spero \[1992\]](#)). Briefly, juvenile specimens of the foraminifera *Orbulina universa* are hand collected using blue-water diving techniques [[Huber et al., 1996](#)]. These specimens are cultured in controlled laboratory conditions at constant temperature under a 12-hr/12-hr day-night cycle in seawater of known composition. Each foraminifera is hand fed every other day with individual *Artemia*. During culture, the juvenile foraminifera develop a spherical terminal chamber, by producing a primary organic sheet (POS) that forms the template for calcite growth on both its inner and outer surfaces. The outer side of the POS is more heavily calcified than the inner side.

The calcite of two specimens was periodically 'labelled' with ^{43}Ca to constrain the origin of calcification, and the position of the POS within the mature skeleton. Specimens of *O. universa* were allowed to precipitate CaCO_3 in natural seawater during the day, and in ^{43}Ca enriched seawater at night to label 'night' calcite over the duration of a multi-day laboratory culture experiment. Because biomineralization is continuous and occurs on both sides of the POS in *O. universa* [[Spero et al., 2015](#)], this night-time ^{43}Ca label generates a predictable number of ^{43}Ca -enriched calcite bands on both sides of the POS.

1.2 Sample Mounting and Polishing

The final calcite spheres of cultured *O. universa* specimens were mounted in Araldite epoxy resin and polished to exposed a cross section using successively finer diamond polishing products (15 - 0.125 μm ; Buehler). The polished surface was sputter-coated with 10nm of gold.

As far as possible, samples were polished until the exposed cross-section was normal to the sphere wall. Oblique sectioning of the sphere by the polished surface will distort chemical patterns along the radial axis. Furthermore, because ToF-SIMS analyses penetrate beneath the sample surface during analysis, laminar features bisecting the surface at an angle may be measured at depth and appear broader than they actually are. The extent of laminar feature distortion due to oblique sample sectioning is calculated from measurements of pre- and post-polishing sphere radius, and post-polishing wall thickness (Fig 1). Based on these calculations, the POS intersected the polished surface at a mean angle of $68\pm 10^\circ$, which corresponds to a sample 'stretch' factor (F , see Fig 1) of 1.1 ± 0.1 (i.e. features appear 1.1 times broader than they actually are in the direction of the growth axis).

1.3 FIB Preparation of APT Specimens

APT specimens were extracted from the polished cross sections using a FEI Helios Nanolab Focused Ion Beam (FIB) instrument (Fig 2). First, a protective layer of Pt was deposited along a strip of calcite $\sim 20 \mu\text{m}$ long and $\sim 2 \mu\text{m}$ wide spanning and parallel to the region thought to contain the POS. The calcite underneath this protected area was excavated from

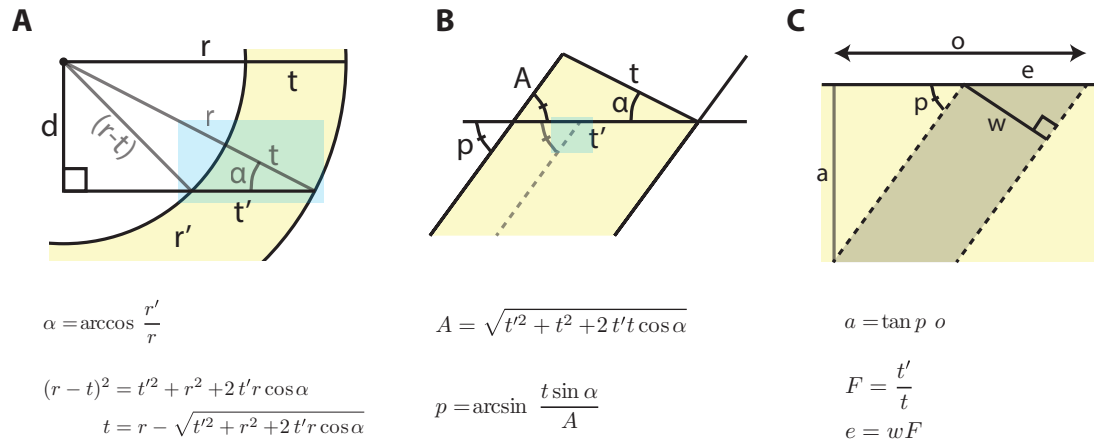


Figure 1: The spherical *O. universa* test may be approximated as two concentric spheres. The angle of sample sectioning may therefore be calculated based on pre- and post- polishing measurements of our specimens. In each diagram, the shaded blue region is enlarged on the right. (A) The radius (r) of the outer sphere is defined by the overall sphere size, and the distance between the inner and outer spheres is the shell thickness (t). Embedding and polishing the shell exposes a plane bisecting these spheres, with a measurable radius (r') and thickness (t'). In the ideal case $r' = r$, and the section will be perfectly normal to the shell wall, but in some cases $r' < r$, and the section will be oblique. Resulting maps will be distorted, and laminae features will be artificially broadened. The extent of this distortion may be calculated as a distortion factor (F), which can be calculated based on our measurements of r , r' and t' . (B) Within the constraints of our analyses (ToF-SIMS penetration depth $< 1 \mu\text{m}$, $F < 1.4$), the intersection of the shell with the resin can be approximated as a linear system, and the angle between the POS and resin (p) calculated. (C) From this, it is possible to calculate the width of a laminae feature of known thickness that bisects the surface (e), and the depth that an analysis must penetrate (a) to make a thin laminae feature appear a given width (o).

the sample surface, yielding a $20 \mu\text{m}$ long strip of calcite, with a triangular cross section and a maximum width of $\sim 2 \mu\text{m}$. The strip was lifted from the sample surface using an Omniprobe micro-manipulator and transferred to a custom sample rotation stage, which allows rotation of the sample along the long axis, to correctly orient features in the APT sample. The rotated sample was then transferred back to the micro-manipulator, and $\sim 2 \mu\text{m}$ sections of the material were cut from the strip and fixed to sharpened Si APT substrate pillars with a Pt weld. The resulting specimens were then sharpened to a fine point using sequentially finer annular milling FIB patterns. As the sample tip became finer, the energy of the FIB beam was reduced to minimize Ga implantation and sample damage. Ga implantation is directly observed in the APT analysis, allowing damaged regions to be excluded during analysis.

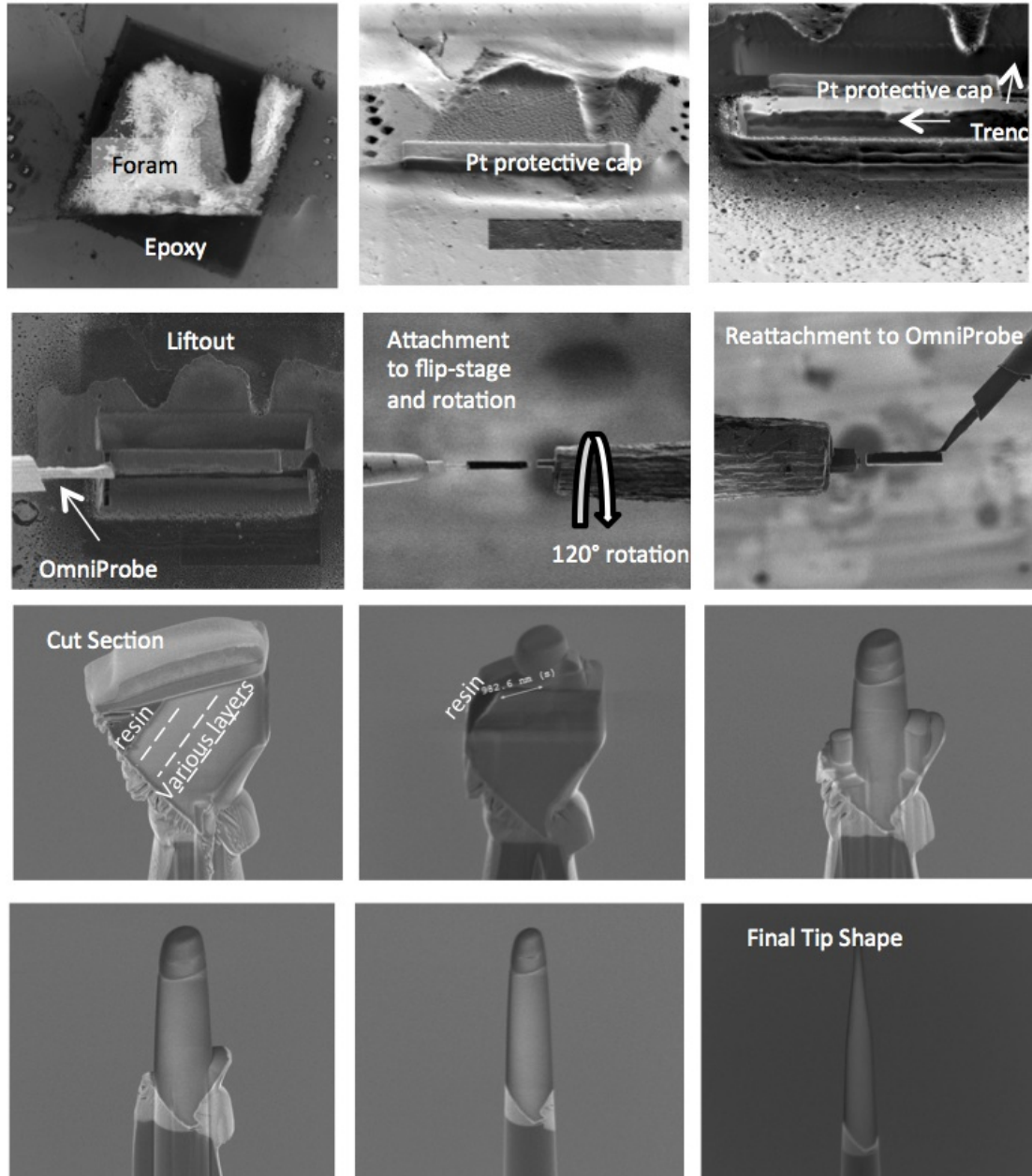


Figure 2: SEM images of FIB sample preparation process starting from a polished section of a foraminiferal shell (roughly 20 microns from inner to outer surfaces) and ending with a roughly 1 micron tall tip ready for APT analysis.

2 Time of Flight Secondary Ionization Mass Spectrometry (ToF-SIMS)

2.1 Data Collection

ToF-SIMS analysis was conducted using a TOF.SIMS5 (IONTOF GmbH, Germany) time-of-flight mass spectrometer. Positive secondary ions were generated through the use of a 25 keV Bi⁺ ion beam with a pulsed beam current of 0.66 pA. The Bi⁺ beam was focused to have a Gaussian full-width-at-half-maximum (FWHM) of approximately 300 nm. The precise beam FWHM was 312 nm, determined from the distance between the 16 – 84th percentiles of the signal change in a linear profile collected across a sharp edge on a 1000 mesh Cu grid standard. This corresponds to a minimum resolvable distance of ~370 nm, following the Rayleigh Criterion of a 26.2% signal decrease between adjacent maxima.

Prior to analysis, the region of interest was pre-sputtered at a 25x25 μm field of view for 60 seconds to clear away the gold coating and reach stable beam conditions. For each chemical map, a total of 3200 measurements were collected from 20x20 μm areas, with a step size of 78 nm (256x256 pixels). Data were collected for all masses between 0-78 amu.

After collection, the maps were analyzed for spatial shift and corrected for translational movement during analysis. The intensity of the beam profile was monitored over the course of each analysis, to ensure stable beam conditions. In most measurements, the beam was unstable for the first half of the analysis (up to scan ~1600). These data were excluded from analyses. The counts at each pixel corresponding to Na, Ca, and Mg collected under stable beam conditions (~1600 maps) were summed to provide chemical maps of the sample. Ca was relatively homogeneous throughout the structure, showing that all observed Mg/Ca and Na/Ca banding are results of changes in the measured Mg and Na intensity.

2.2 Data Calibration

ToF-SIMS Na/Ca and Mg/Ca were calibrated using six carbonate standards (Table 1). Previous measurements of these standards by isotope dilution ICP-MS [Gabitov et al., 2013] revealed a degree of Mg and Sr heterogeneity between grains. We therefore calibrated our ToF-SIMS profiles against Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS) analyses of Mg/Ca and Na/Ca taken from the same grain surfaces measured by ToF-SIMS. Analyses were performed using a Teledyne Photon Machines Analyte G2 193 nm excimer laser with a HelEx dual-volume laser ablation cell coupled to an Agilent 7700x quadrupole-ICP-MS. The LA-ICP-MS measured Mg/Ca and Sr/Ca of standards were all within range of published isotope-dilution ICP-MS measurements [Gabitov et al., 2013].

2.3 Data Processing

Matrices containing raw ToF-SIMS counts of Ca, Na and Mg were imported into python for analysis using the numpy and scipy packages. Masks were created to remove counts originating from the mounting resin by smoothing the Ca image with a 3 pixel Gaussian filter, and manually specifying a minimum count threshold based on the Ca count histogram

Table 1: The names, abbreviations and M/Ca in mmol/mol of the six standards used in this study, as determined by LA-ICP-MS.

Standard	Mg/Ca (mmol/mol)	Na/Ca (mmol/mol)
Blue Calcite (BCC)	4.23 ± 0.07	0.015 ± 0.003
UC Berkeley Calcite 135 (CAL-135)	0.108 ± 0.008	0.009 ± 0.001
Hebrew University Aragonite (HUJ)	0.0463 ± 0.0004	1.08 ± 0.02
LAS-20 Calcite (LAS-20)	43.00 ± 0.53	0.48 ± 0.03
Rensselaer Polytechnic Calcite (RPI)	0.402 ± 0.006	0.011 ± 0.002
UC Irvine Calcite (UCI)	3.52 ± 0.08	0.012 ± 0.002

to separated ‘calcite’ and ‘resin’ regions within the map. Next Mg/Ca and Na/Ca ratios were calculated from the masked matrices, and calibrated against two or more of the carbonate standards (Table 1), measured during the same ToF-SIMS analysis session. The mean Na/Ca and Mg/Ca ratios of entire analysis regions was within error of LA-ICP-MS measurements of foraminifera cultured under similar conditions (2.6 ± 0.2 mmol/mol Na/Ca, and 6 ± 1 mmol/mol Mg/Ca), demonstrating the efficacy of our calibration.

Profiles of $^{43}\text{Ca}/\text{Ca}$, Mg/Ca and Na/Ca were extracted along the growth axis of the shell, normal to the laminar compositional patterns. This was achieved by rotating the ToF-SIMS matrices using linear interpolation to re-grid the data, and extracting sub-matrices from an region of interest (ROI) within the rotated image. This sub-matrix was then converted to a concentration profile by calculating the per-row mean normal to the growth axis.

2.4 POS Identification

Several lines of evidence identify the location of the POS within our ToF-SIMS maps:

1. The introduced periodic ^{43}Ca ‘label’ present in the calcite which symmetrically brackets the origin of calcification (Section 1.1).
2. Skeletal Mg/Ca increases naturally each night [Eggins et al., 2004, Spero et al., 2015], offering a similar, though less controlled periodic marker.
3. ToF-SIMS maps of a calcite spine, which terminates at the POS.

In the two specimens containing the ^{43}Ca label, the origin of calcification can be constrained to between the innermost ^{43}Ca bands (Fig 3A). In the specimen analyzed by APT, this corresponds to a $\sim 1.17 \mu\text{m}$ wide band. At approximately the center of this band, we observe a single, distinct Na/Ca maximum, bracketed by double Mg/Ca maxima. This double Mg/Ca maxima was observed in $\sim 33\%$ of specimens (Fig 3 A & B), as opposed to the remainder of specimens which contained a single distinct Mg maximum coincident with the Na maximum (Fig 3C). This feature can explained if chamber formation is initiated in the late afternoon, causing two distinct layers of high-Mg ‘night’ calcite [Eggins et al., 2004, Spero et al., 2015] to be deposited on either side of the POS in close proximity to the origin of calcification. Because calcification occurs on both sides of the POS, the origin of

calcification and the position of the POS may be constrained to the ~ 600 nm between these bracketing Mg/Ca maxima. These maxima are located ~ 1 FWHM on either side of the POS, and will therefore have minimal contribution to the POS Mg signal. Given the analysis spot (FWHM = 312 nm) and step (78 nm) sizes, the majority of analysis spots between these adjacent Mg/Ca maxima will measure the POS. Thus, in the specimen analyzed by APT, we are able to constrain the position of the POS to within the resolution limits of the instrument. In both ^{43}Ca labeled specimens, the innermost ^{43}Ca bands contained simultaneous Mg/Ca and Na/Ca maxima. These elements did not coincide systematically anywhere else in the cross sections outside the origin of calcification, and appear to be uniquely associated with the position of the POS.

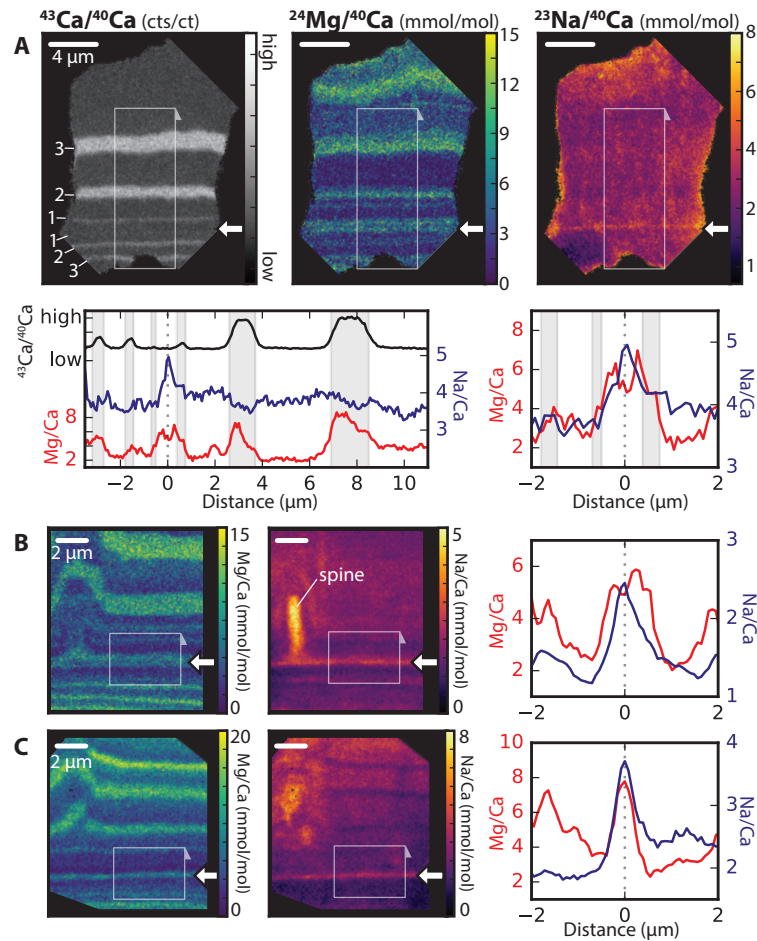


Figure 3: ToF-SIMS maps reveal the distinctive chemistry of the POS. Maps of $^{43}\text{Ca}/\text{Ca}$, Mg/Ca and Na/Ca in three *O. universa* specimens, and the chemistry of the POS. The POS position is marked by a white arrow on the right side of all images, and a dotted vertical line in all profiles. Profiles are extracted from white boxes on the images, with the direction of the distance axis marked by an arrow. Profiles on the right highlight the POS-associated Na/Ca and Mg/Ca enrichment. (A) Maps of *O. universa* specimen analyzed by APT, containing a $^{43}\text{Ca}/\text{Ca}$ 'night' calcite label, allowing the identification of the POS. (B) A specimen containing a Na-rich spine originating at the POS. Both (A) and (B) have double Mg maxima either side of the Na maximum, due to the POS being enveloped by two high-Mg bands. This was present in $\sim 1/3$ of specimens analyzed, and a more typical POS-signal is shown in (C).

In the 17 specimens *without* a ^{43}Ca label, the location of the POS was identified using the naturally occurring nightly Mg/Ca maxima throughout the shell, which offer a less precise estimate of POS location (Fig 3B&C). In all specimens the region identified using Mg/Ca periodicity contained a simultaneous Mg/Ca and Na/Ca maxima, similar to that observed associated with the POS in the ^{43}Ca labeled specimens. This suggests that the simultaneous Mg/Ca and Na/Ca maxima observed in all specimens is uniquely associated with the origin of calcification, and the position of the POS.

Finally, the location of the POS is independently confirmed by ToF-SIMS maps which contain the base of a 'spine', a calcite structure which is known to originate at the POS and protrude through the shell wall [Spero, 1988]. Our ToF-SIMS maps reveal that these spines are significantly enriched in Na, and that they originate from the POS-associated Na and Mg enriched band (Fig 3B).

In combination, these localization techniques point to a single high Na and Mg enriched band at the origin point of calcification, associated with the POS. Coincident Na and Mg maxima in *O. universa* may provide a convenient intra-skeletal tracer for POS location, which may be applied in future studies targeting the POS.

Our maps also reveal systematic 'banded' Na/Ca variability (Fig 3 B & C), which is distinct from the POS signal. The magnitude of this variability is significantly less than the POS Na/Ca maximum. Away from the POS, the Na/Ca maxima rarely coincides with Mg/Ca maxima, and Na/Ca often appears anti-correlated with the well-characterized distribution of Mg/Ca found throughout the skeleton [Fig 3; Eggins et al., 2003, Spero et al., 2015] Because X-ray analyses of *O. universa* have revealed that Mg away from the POS is not hosted in organic components [Branson et al., 2013], this suggests that non-POS chemical heterogeneity is unlikely to be driven by organic material.

2.5 POS Measurement

The mean Na/Ca and Mg/Ca signal in concentration profiles normal to the POS were fit with a combination of a Gaussian function and a linear background to precisely determine the width, amplitude and position of the POS-associated signal (Fig 4). The Na and Mg maxima were found to have a mean FWHM of 673 ± 272 nm. Estimates of POS width from previous studies, and our 'scaling-up' model suggest it is ~ 100 nm wide. Because of oblique sample sectioning (Section 1.2), the sub-surface analysis of material by the ToF-SIMS beam could cause this broadening. However, to measure a ~ 100 nm feature as a ~ 670 nm feature, the ToF-SIMS beam would have to penetrate $2.2 \mu\text{m}$ into the sample surface, which is double the maximum estimated analysis depth of $\sim 1 \mu\text{m}$. It is therefore unlikely that this broadening can be explained by the geometry of sample sectioning.

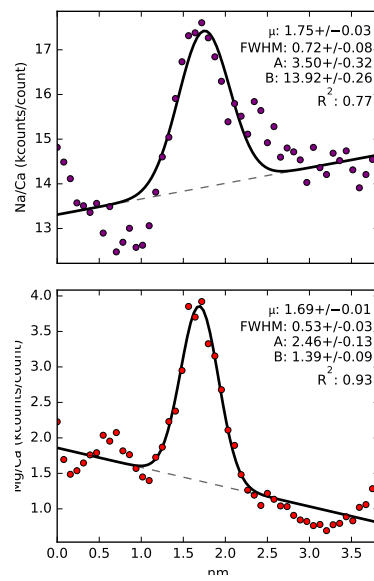


Figure 4: ToF-SIMS measured Na/Ca and Mg/Ca maxima associated with the POS, fitted with a linear background and a Gaussian peak. Statistics noted on the plots: μ = peak position, FWHM = peak Full Width at Half Maximum, A = height of the Gaussian at μ , B = the background at μ , and the R^2 of the fit.

3 Atom Probe Tomography (APT)

3.1 Specimen Optimization

The thermal insulating properties of calcite slow the dissipation of heat from the laser pulses, which can lead to heterogeneous evaporation, changes in complex ion formation, and thermal artifacts in the resulting ToF spectra as ionization harmonics are generated by reflected energy within the sample. These complex effects are determined by subtleties of sample geometry and are challenging to correct for post-analysis. A broad tip with a high shank angle provided the 'cleanest' ToF spectra, because it provided maximal area for heat to dissipate down the sample (Fig 2). This tip geometry also minimized the rate of catastrophic fracture in the tips.

In addition, a variety of tip geometries and sample coatings were investigated in attempts to optimize run success. The electrical insulating properties of calcite make the sample tips prone to fracture under a high applied field, as the hydrostatic stresses induced by the field exceed the strength of the sample tip. Conductive sample coatings (Pt, Cr) were investigated, but differences between the ionization fields between coating and specimen led to the evolution of tip topography that introduced complex artifacts in the samples which rendered the data unusable (Fig 5). Future research to identify a sample coating with a similar evaporation field to calcite has the potential to significantly improve the likelihood of successful analysis. In our study, the highest quality data were collected from uncoated samples. Sample fracture at the start of analysis was minimized by manually ramping up sample potential voltage very slowly, while constantly monitoring ionization rates and patterns.

3.2 Data Collection

Sample tips were analyzed using a Cameca Local Electrode Atom Probe (LEAP 4000X-HR) running in pulsed laser mode (at 44 K, with a 50 pJ 355 nm UV laser pulsed at 160-200 Hz). Laser pulse rate was varied throughout the run to minimize thermal artifacts in the sample spectrum. Specimen voltage was increased manually to 4 keV, or until sufficient ion yield was obtained to allow fine sample alignment relative to the detector and local electrode. After alignment, voltage was automatically controlled to maintain an ion yield of 0.2%. Typically the voltage increased throughout the run, but did not exceed 7 keV. On average, ions were excited every 1033 laser pulses, with 1.13 ions per excitation (19.5 % multi-hits). The mass resolving power is 1270 for $^{40}\text{Ca}^{2+}$, and 365 for Na^+ . Mass resolving power is defined as $M/\Delta M$ using peak width at full width and half maximum height (FWHM) to measure ΔM .

Approximately 10% of ~30 specimens analyzed yielded usable datasets ($> 1 \times 10^6$ counts), although explosive ionization events driven by structural heterogeneities reduced our success rate to ~5%. In contrast, ~90% of samples fractured catastrophically in the initial stages of analysis.

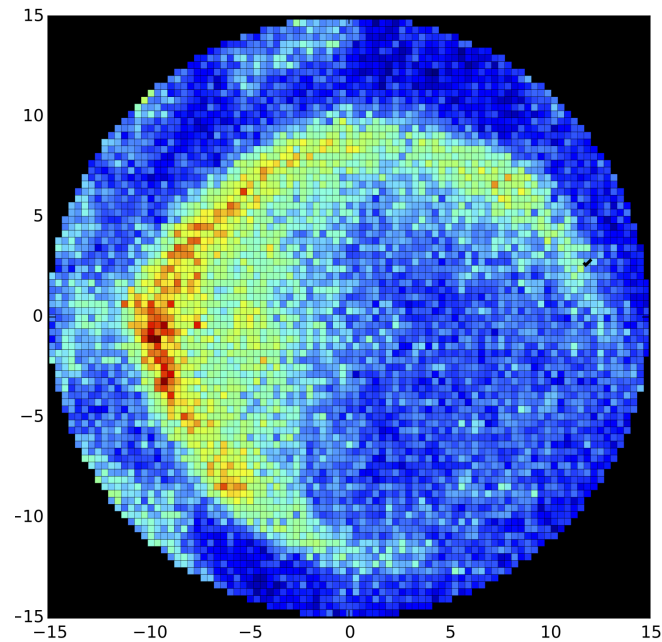


Figure 5: Total detector hits during the analysis of a Cr coated specimen, demonstrating the complications associated with some of the coatings we experimented with. The ring represents anomalously high counts, associated with the deflection of ions by topography at the Cr/calcite boundary.

3.3 Data Processing

Raw LEAP data were processed using a combination of Cameca's IVASTM software and python scripts in the IPython environment [Perez and Granger, 2007, Section 5].

3.3.1 Spectral Processing

In IVAS, ToF peaks were corrected for the hemispherical shape of the peak tip by minimizing their full width at half maximum, within the constraints of a physical instrument model. The ToF spectra were then scaled to known peak positions to account for variations in sample-detector distance.

Peaks in the ToF-SIMS spectra were approximately identified by hand, then precisely characterized in python using an asymmetric pseudo-voigt function, adapted from Stancik

and Brauns [2008]:

$$\gamma_x(x) = \frac{2\gamma_0}{1 + \exp[a(x - x_0)]} \quad (1)$$

$$G(x) = \frac{A}{\gamma_x} \sqrt{\frac{4 \ln 2}{\pi}} \exp \left[-4 \ln 2 \left(\frac{x - x_0}{\gamma_x} \right)^2 \right] \quad (2)$$

$$L(x) = \frac{2A/\pi\gamma_x}{1 + 4[(x - x_0/\gamma_x)]^2} \quad (3)$$

$$y(x) = fL(x) + (1 - f)G(x) \quad (4)$$

Where γ denotes peak's full width at half max (FWHM), and varies as a function of x to impart asymmetry to the peak. The degree of asymmetry is determined by the variable a . The balance between Lorentzian ($L(x)$) and Gaussian ($G(x)$) components is determined by the coefficient f . The instantaneous FWHM (γ_0) determines the overall peak width.

Peak and background windows were chosen based on these fits, and used for all subsequent analysis of bulk and proxygram compositions. The 'on-peak' window was specified as 1 FWHM either side of the peak center, and two 1 FWHM wide 'background' windows were placed either side of the peak 3 FWHM away from the peak center. Background and peak windows were examined manually before proxygram creation, and the position of background regions were adjusted to avoid adjacent peaks and significant peak tails. Where the peak tail was unavoidable (e.g. $^{40}\text{Ca}^{2+}$), both background regions were moved to before the peak.

These peak assignments were used to background correct ToF spectra and 'proxygram' composition profiles extracted from IVAS. First, each peak extracted in the proxygram was background corrected, assuming that the background was linear in the peak region:

$$C(d) = C_{pk} - C_{bg} * \frac{n_{pk}}{n_{bg}} \quad (5)$$

where n is the number of bins of peak and background counts captured by their respective windows. The error of the background based on the standard deviation of the background region is taken into account during background subtraction. Background corrected counts are converted to atom counts by decomposing complex ions, based on the compositions assigned to the peaks (Section 5.3.2).

ToF peaks were assigned compositions based on the most likely combination of constituent ions known to be present in the sample (Table 2 and Fig 6). The bio-carbonate samples yielded numerous complex and multiply charged ions, and it was not possible to identify all peaks. Of the total 45 ToF peaks, six could have been produced by multiple likely compositions, so were excluded from analysis. These six unidentified peaks accounted for 1.45% of all on-peak ToF counts, and 98.55% of on-peak counts were assigned ionic identities.

Table 2: All peaks observed in the ToF spectra, and their ionic identities

Da	Composition	Da	Composition
1.0	H ⁺	27.9	¹² C ¹⁶ O ⁺
2.0	H ₂ ⁺	29.0	¹² C ¹⁶ OH ⁺
3.0	H ₃ ⁺	30.0	¹² C ¹⁶ O ₃ ²⁺
6.0	¹² C ²⁺	30.5	<i>Unknown</i>
6.5	¹² CH ²⁺ / (¹³ C ²⁺)	31.0	³¹ P ⁺
7.0	¹⁴ N ²⁺	32.0	¹⁶ O ₂ ⁺
12.0	¹² C ⁺ / ²⁴ Mg ²⁺	36.0	⁴⁰ Ca ¹⁶ O ₂ ²⁺
12.5	²⁵ Mg ²⁺	37.0	<i>Unknown</i>
13.0	¹² CH ⁺ / ²⁶ Mg ²⁺	40.0	⁴⁰ Ca ⁺
14.0	¹² CH ₂ ⁺ / ¹⁴ N ⁺	42.0	⁴² Ca ⁺
15.0	¹² CH ₃ ⁺ / ¹⁴ NH ⁺	43.9	¹² C ¹⁶ O ₂ ⁺
16.0	¹⁶ O ⁺ / ¹⁴ NH ₂ ⁺	44.9	<i>Unknown</i>
17.0	¹⁶ OH ⁺ / ¹⁴ NH ₃ ⁺	48.0	⁴⁰ Ca ¹⁶ O ₂ ²⁺
18.0	¹⁶ OH ₂ ⁺ / ¹⁴ NH ₄ ⁺	55.9	⁴⁰ Ca ¹⁶ O ⁺
18.5	<i>Unknown</i>	57.0	⁴⁰ Ca ¹⁶ OH ⁺
19.0	¹⁶ OH ₃ ⁺	59.0	<i>Unknown</i>
20.0	⁴⁰ Ca ²⁺	59.9	¹² C ¹⁶ O ₃ ⁺
21.0	⁴² Ca ²⁺	61.9	<i>Unknown</i>
22.0	⁴⁴ Ca ²⁺	68.8	Ga ⁺
23.0	⁴⁶ Ca ²⁺ / ²³ Na ⁺	69.9	⁴⁰ Ca ¹² C ¹⁶ O ₃ ²⁺
24.0	⁴⁸ Ca ²⁺ / ²⁴ Mg ⁺	71.9	⁴⁰ Ca ¹⁶ O ₂ ⁺
25.0	²⁵ Mg ⁺	99.9	⁴⁰ Ca ¹² C ¹⁶ O ₃ ⁺
26.0	²⁶ Mg ⁺		

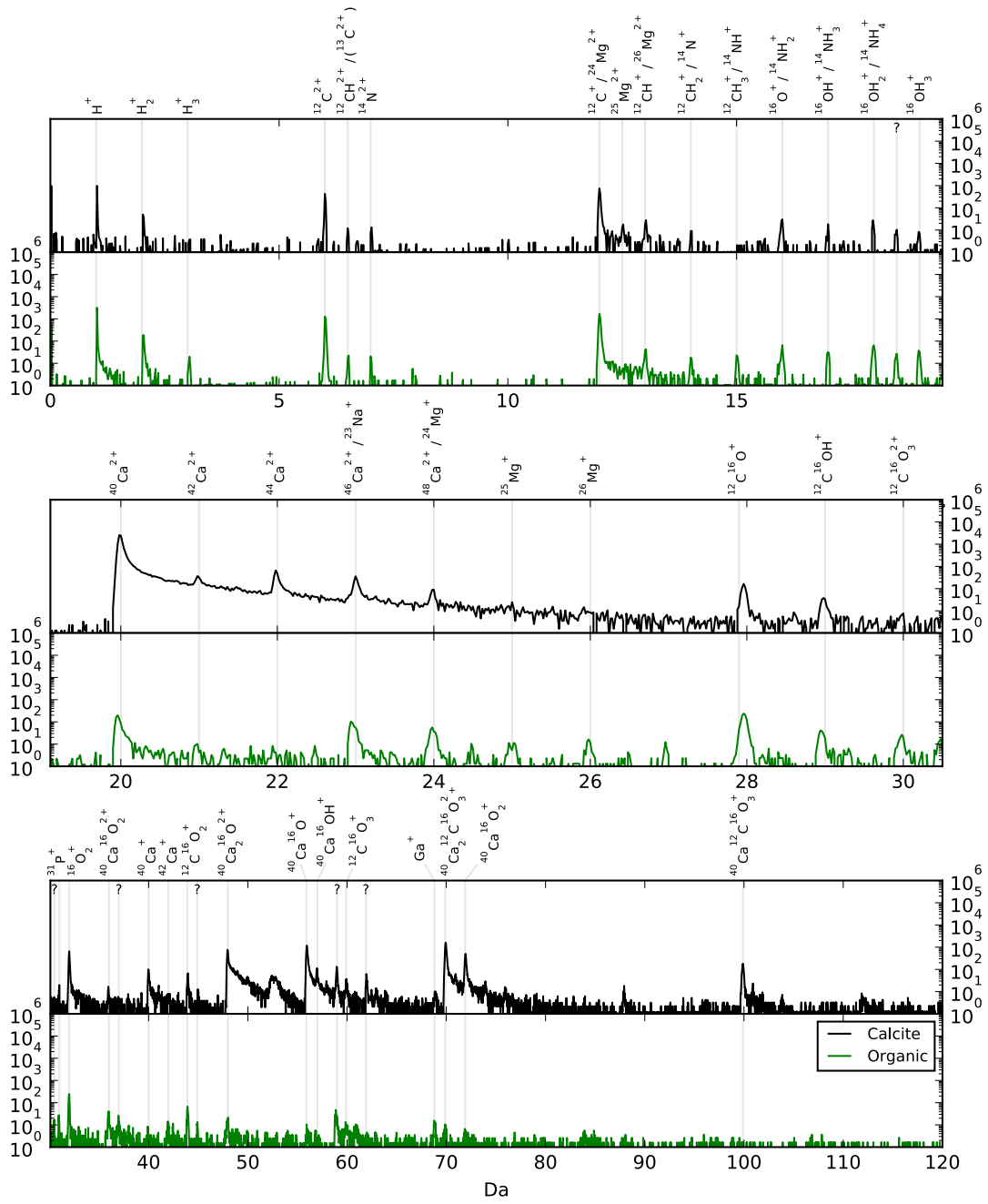


Figure 6: ToF-SIMS spectra from calcite and organic layers, with peak identities (Table 2) marked.

3.3.2 Na and Mg calculation

The main Na and Mg peaks suffer from interferences with major element peaks in the Time-of-Flight spectrum (Table 3). These interferences may be accounted for by either subtracting calculated interference counts in the case of Na, or by extrapolation from counts of the minor isotope in the case of Mg (Fig 7). For a detailed evaluation of interference correction, see Section 5.3.3.

Table 3: Possible ions of Na and Mg and their Interferences.

Da	Isotope of Interest	Relative Abundance	Interference
23.0	$^{23}\text{Na}^+$	100%	$^{46}\text{Ca}^{2+}$
12.0	$^{24}\text{Mg}^{2+}$	79%	$^{12}\text{C}^+$
12.5	$^{25}\text{Mg}^{2+}$	10%	Unknown
13.0	$^{26}\text{Mg}^{2+}$	11%	$^{12}\text{CH}^+$
24.0	$^{24}\text{Mg}^+$	79%	$^{48}\text{Ca}^{2+}$
25.0	$^{25}\text{Mg}^+$	10%	Unknown
26.0	$^{26}\text{Mg}^+$	11%	Unknown

Sodium does not have any interference-free peaks, so must be calculated by subtracting calculated interference counts. Interference counts for the 23 Da peak are calculated by estimating $^{46}\text{Ca}^{2+}$ counts from $^{40}\text{Ca}^{2+}$ counts and the natural abundance of Ca isotopes:

$$^{23}\text{Na}_c^+ = ^{23}\text{Da}_c - ^{46}\text{Ca}_c^{2+} \quad (6)$$

$$^{46}\text{Ca}_c^{2+} = ^{40}\text{Ca}_c^{2+} \times A^{46}\text{Ca}/A^{40}\text{Ca} \quad (7)$$

Where the relative abundance (A) of $^{46}\text{Ca}^{2+}$ and $^{40}\text{Ca}^{2+}$ are 0.004% and 96.941%, respectively.

Total Mg counts may be extrapolated from both ^{25}Mg and ^{26}Mg counts, which have no known interferences with major sample constituents. In the case of singly-charged Mg, both $^{25}\text{Mg}^+$ and $^{26}\text{Mg}^+$ are evaluated. It is possible that unknown complex ion interferences may contribute to these peaks, so the lesser of the two peaks is used to calculate total Mg counts, following:

$$^T\text{Mg}_c^+ = ^{26}\text{Mg}_c^+ \times 100/A^{26}\text{Mg} \quad (8)$$

Where the $A^{26}\text{Mg}$ is the percentage abundance of ^{26}Mg (11%). For doubly charged Mg, significant interference peaks are present at 12 and 13 Da (Table 3), so $^{25}\text{Mg}^{2+}$ counts at 12.5 Da are used to estimate total Mg^{2+} counts:

$$^T\text{Mg}_c^{2+} = ^{25}\text{Mg}_c^{2+} \times 100/A^{25}\text{Mg} \quad (9)$$

Where the $A^{26}\text{Mg}$ is the percentage abundance of ^{25}Mg (10%). Single and double charged Mg counts may then be considered independently, or summed to calculate total Mg counts.

For Na, the same correction procedure is applied to both bulk APT spectra for calcite and organic components, and during proxygram analysis. For Mg, the sum of $^{25}\text{Mg}^{2+}$ and

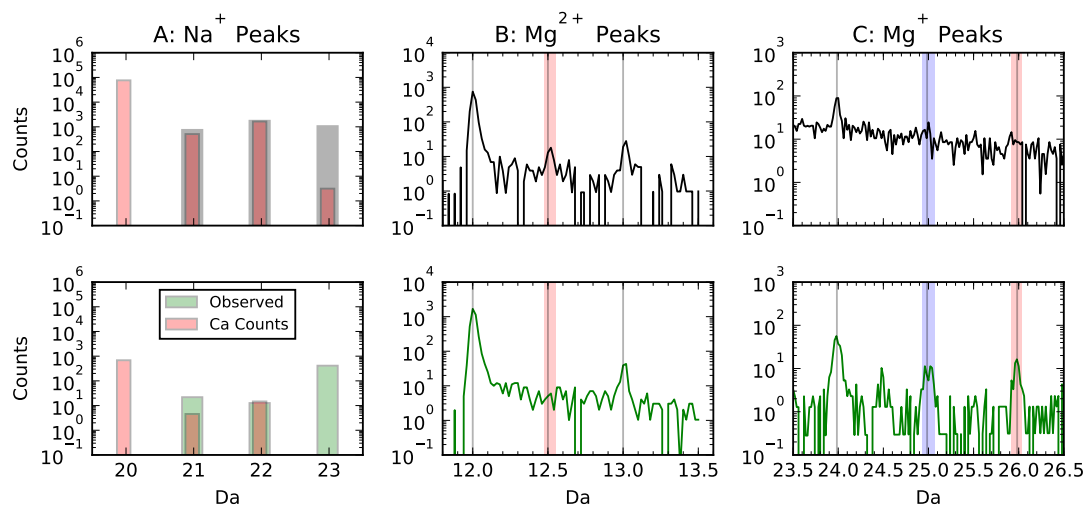


Figure 7: The correction methods used to deconvolve Na and Mg counts from interference peaks. (A) Predicted Ca counts for the 21, 22 and 23 Da peaks are based on measured $^{40}\text{Ca}^{2+}$ counts and the natural abundance of Ca isotopes. Na is calculated as the difference between the measured 23 Da peak, and the predicted $^{46}\text{Ca}^{2+}$ counts. (B & C) Total Mg counts are calculated from minor $^{25}\text{Mg}^{2+}$ and $^{\text{Mg}}\text{Ca}^{+}$ (highlighted in red). In (C), where both $^{25}\text{Mg}^{+}$ and $^{26}\text{Mg}^{+}$ peaks are evident, the smaller of the two is used, as the larger is more likely to be affected by unconstrained interferences. For a full evaluation of Mg estimation methods, see Section 5.3.3.

$^{26}\text{Mg}^{+}$ are used when dealing with whole-region mass spectra. However, only the $^{26}\text{Mg}^{+}$ peak is used in proxygram calculations. This is because the $^{25}\text{Mg}^{2+}$ peak at 12.5 Da is negligible in the spectrum extracted from the entire organic region (Fig 7), so considering this peak in proxygrams, where each bin considers 800 counts, would contribute more noise to the proxygram than Mg counts (Section 5.5).

3.3.3 3D Reconstruction

The entire data volume was reconstructed to a 3D point cloud, constraining the tip shape and total volume analyzed to physical measurements of sample tip. The tip shape and volume was determined from measurements of pre- and post-analysis SEM images of the tip, corrected for SEM view angle (Fig 8). This provided a tip radius estimate which evolves with z , and defines the reconstruction shape (Fig 9). The volume of the reconstruction was further altered by tuning the mean ion size and the image scaling factor so the reconstructed volume matched the volume of material removed during LEAP analysis.

The image compression factor is an empirical parameter that is used to account for the compression of electric field lines during analysis of a needle shaped specimen. The mean ion size is used to determine the spatial arrangement of ions in the reconstruction. By default, the ionic radius of the most common ion is used (in this case, Ca). However, the high proportion of complex ions from our samples rendered this approach inaccurate, so this parameter was increased to expand the reconstructed volume to match the volume of material removed by the LEAP. Final reconstruction parameters can be seen in Table 4. The

specimen radius parameter is used to determine the z evolution of atom position assignment: xy are measured by the detector, and z position is increased incrementally during position assignment, based on ion number.

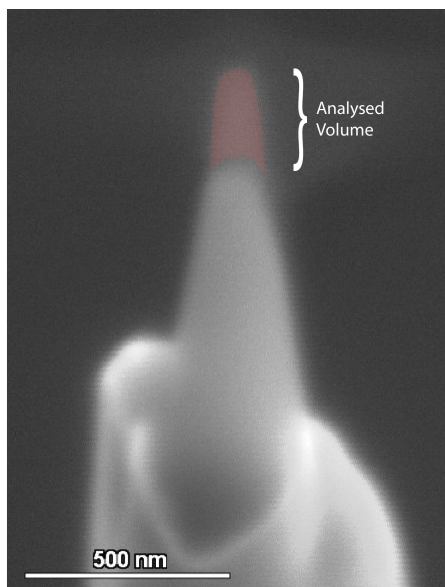


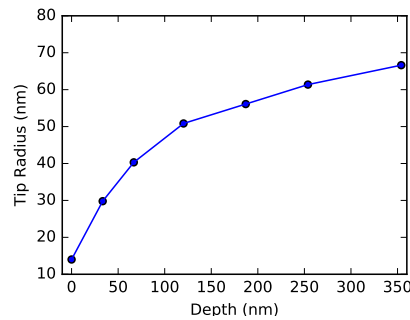
Figure 8: Overlaid pre- and post-analysis SEM images of a sample tip, with the analyzed region highlighted in red.

3.3.4 Interface Identification

Two materials are evident in our reconstructed data, as identified by differences in: (1) the abundance of Ca, C and H (2) changes in complex ions, particularly in C-bearing (3) the evaporation field. In calcite, C tends to ionize as CO_x^{n+} and $\text{Ca}_x\text{CO}_y^{n+}$ complexes, reflecting the CaCO_3 origin of these ions. In contrast, C^{n+} and CH_x^{n+} ions are more prevalent in organic material, reflecting the typical bonding environment of C in organic compounds. This C ionization pattern agrees with the findings of previous APT analyses of organic materials [Gordon and Joester, 2011]. Calcite is further characterized by abundant Ca and a relatively low evaporation field, while the organic is characterized by low Ca, high elemental C and a relatively high evaporation field. The interface between the materials is defined as the 50% Ca isosurface through a $1 \times 1 \times 1.5$ nm xyz sample mesh, after complex ion decomposition.

Table 4: Instrumental and reconstruction parameters for the specimen presented in the main manuscript.

	Parameter	Value
Instrument Parameters		
	k-factor	3.30
	Detector Efficiency	0.36
	Image Compression Factor	1.0
Materials Parameters		
	Primary Element	Ca
	Evaporation Field	18.0 V/nm
	Atomic volume	0.0435 nm ³
Specimen Parameters		
	Initial Tip Radius	14.0 nm
	Radius Evolution	see Fig 9
	Estimate Shank Half Angle	28.5 degrees

**Figure 9:** The evolution of tip radius used in the reconstruction, as measured from Fig 8. The depth axis is corrected for an SEM view angle of 56°.

3.3.5 Proxygram extraction and processing

Proxygrams are concentration profiles created normal to a defined interface. In its raw form, it provides the ion counts of all regions of the ToF spectrum specified in a ‘range’ file, which contained ‘on-peak’ and adjacent ‘background’ regions for each ToF peak, as determined by a peak fitting algorithm (Section 3.3.1).

Proxygrams were exported with a fine bin size (0.01 nm), and re-binned such that:

$$\sum_d^{d+w} A = n \quad (10)$$

Where bin width (w) at a given position in the proxygram (d) varies to maintain a constant total atom (A) count (n). This reveals relative atomic composition across the proxygram, normalizing for differences in ion yield between the materials.

For a more practical, in-depth description of this procedure, see Section 5.

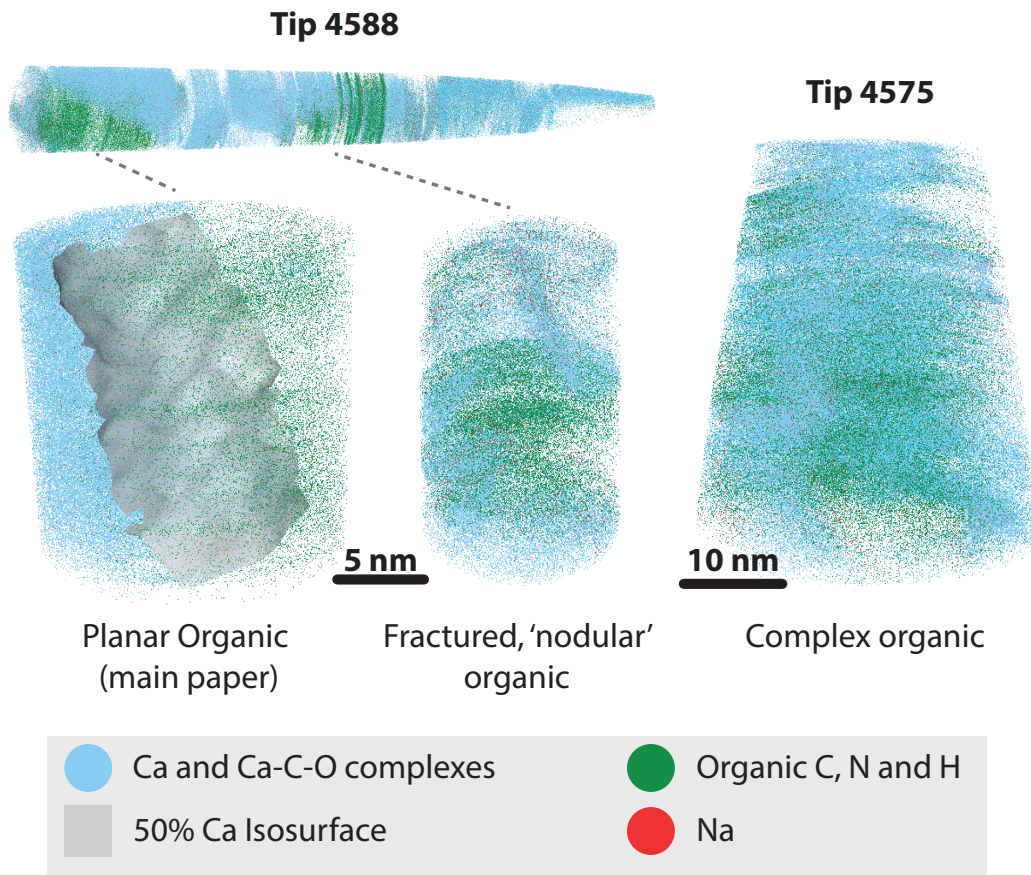


Figure 10: Three organic regions were captured across two APT specimens.

3.4 Secondary Organic Regions

In addition to the large, planar organic-calcite interface region, we observe two other 'nodular' calcite-embedded organic regions in separate APT specimens (Fig 10). Although these nodular regions were too small for robust compositional analysis, they may be non-laminar extensions of the POS, or previously undocumented non-POS mineral-associated organics (Main Paper Fig 1C). These non-POS organics suggest that in *O. universa*, organic-mineral interactions may play a role in biomineralization beyond that of the POS.

4 Scale-Up Model

4.1 Context

We have measured the atom-scale composition of calcite and an embedded organic using APT, and sub-micron-scale skeletal chemistry using ToF-SIMS. We have observed that both the organic observed by APT, and the region associated with the position of a known organic layer in the skeleton measured by ToF-SIMS are both enriched in Mg and Na, compared to the adjacent calcite. This qualitative similarity suggests that the signal observed by ToF-SIMS may be caused by the same organic that we have captured by APT. We seek to determine whether the organic measured by APT can quantitatively account for the Na and Mg enrichment observed in ToF-SIMS. However, these data cannot be directly compared because:

1. The spatial scale of the measurement are not comparable.
2. It is not possible to calibrate the APT measurements.
3. Differences in ionization mechanics do not allow direct comparison of raw counts.

To overcome these barriers, we use a quantitative model to simulate the magnitude of the compositional signal we would expect if we observed a calcite-embedded organic layer by ToF-SIMS, using the compositions of the calcite and organic components measured by APT.

4.2 Hypothesis

Our model tests the hypothesis that:

The Ca, Mg and Na counts measured in calcite and organic materials by APT match the POS-associated Mg/Ca and Na/Ca enrichment seen in ToF-SIMS.

If this hypothesis is true, it follows that the POS-associated chemical signal observed by ToF-SIMS originates from the same organic material observed by APT.

Testing this hypothesis with our model is not immediately straightforward. Because the two techniques integrate over different spatial scales, the magnitude of both Mg and Na signals will be different between APT and ToF-SIMS. However, the relative ratio of Mg to Na within one method must match the ratio of Mg to Na in the other method. To accept our hypothesis we must show that the $(\text{Na}/\text{Ca})/(\text{Mg}/\text{Ca})$ in the APT-measured organic is the same as in the ToF-SIMS signal. For the reasons outlined above it is not possible to directly compare these datasets, so we must use a quantitative model.

We may accept our hypothesis if we are able to simulate *both* the Mg/Ca and Na/Ca enrichments observed in ToF-SIMS with a single model parameterization. If matching Na/Ca and Mg/Ca enrichments requires two separate model parameterizations, the hypothesis must be rejected.

4.3 Parameterization

The model involves two steps:

1. Simulate the proportion of a ToF-SIMS analysis beam that would illuminate organic material as it scanned across a calcite-embedded organic layer.
2. Use an end-member mixing model to translate the 'proportion organic material' to a compositional signal, based on APT-measured Ca, Mg and Na counts in the two materials.

4.3.1 Simulated Analysis Beam

We simulate a ToF-SIMS analysis beam scanning across a calcite-embedded organic layer of specified thickness. In mathematical terms, this is reduced to 2 dimensions, and accomplished by convolving a Gaussian function with a 'tophat' function to calculate the proportion of the ToF-SIMS analysis beam that measures the organic and calcite materials (Figure 11).

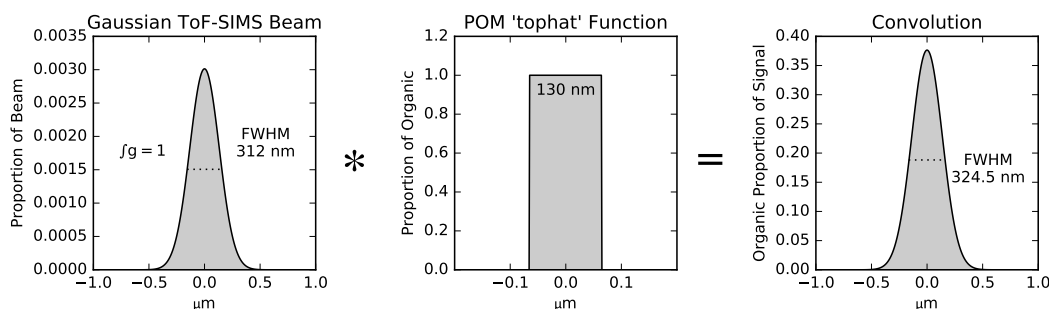


Figure 11: The calculation of the organic signal intensity and shape by convolution. (A) The Gaussian ToF-SIMS beam shape with a FWHM of 312 nm and an integral of 1. (B) The tophat function describing the organic layer, shown here with a total width of 130 nm. (C) The convolution of the gaussian and tophat functions, whose intensity is equal to the proportion of the ToF-SIMS signal that comes from the organic layer, and whose width is a function of both organic layer width and beam FWHM.

This convolution is performed numerically using `numpy.convolve`, which produces a signal whose intensity corresponds to the proportion of organic material measured by the analysis beam (see Appendix 4.7: Convolution Calculation).

4.3.2 End-Member Mixing Model

Once the proportion of organic illuminated by the simulated analysis beam is determined, this is applied to a simple end-member mixing model. The M/Ca ratio 'seen' by the simulated analysis beam is defined by the total counts of M and Ca from each material:

$$\frac{M}{Ca_{mix}} = R_{mix} = \frac{M_{nO} + M_{nC}}{Ca_{nO} + Ca_{nC}} \quad (11)$$

Where n is the number of counts, the superscript denotes the ion ($M = \text{non-Ca}$), the subscript denotes Calcite or Organic, and $\frac{M}{Ca}$ is expressed as R . From this:

$$R_{mix}^{Ca_{nO}} - M_{nO} = M_{nC} - R_{mix}^{Ca_{nC}} \quad (12)$$

And, dividing through by Ca_{nO} :

$$Ca_{nO}(R_{mix} - R_O) = Ca_{nC}(R_C - R_{mix}) \quad (13)$$

$$\frac{Ca_{nO}}{Ca_{nC}} = \frac{(R_C - R_{mix})}{(R_{mix} - R_O)} \quad (14)$$

$$\left(\frac{Ca_{nO}}{Ca_{nC}}\right) R_{mix} - \left(\frac{Ca_{nO}}{Ca_{nC}}\right) R_O = R_C - R_{mix} \quad (15)$$

$$R_{mix} \left(\frac{Ca_{nO}}{Ca_{nC}} + 1\right) = R_C + \left(\frac{Ca_{nO}}{Ca_{nC}}\right) R_O \quad (16)$$

$$R_{mix} = \frac{R_C + \left(\frac{Ca_{nO}}{Ca_{nC}}\right) R_O}{\left(\frac{Ca_{nO}}{Ca_{nC}} + 1\right)} \quad (17)$$

With knowledge of the relative Ca content of the calcite and organic, and the proportion (p) of each material 'seen' by the simulated analysis beam, $\frac{Ca_{nO}}{Ca_{nC}}$ may be calculated as:

$$\frac{Ca_{nO}}{Ca_{nC}} = \frac{[Ca]_O p_O}{[Ca]_C p_C} \quad (18)$$

Where the proportions of calcite (p_C) and organic material (p_O) are calculated by convolution (Section 4.3.1) or via a geometric model describing bulk calcification (Section 4.6). Thus, the convolved organic layer signal may be converted to a compositional Mg/Ca or Na/Ca signal following:

$$R_{mix} = \frac{R_C + \left(\frac{[Ca]_O p_O}{[Ca]_C p_C}\right) R_O}{\left(\frac{[Ca]_O p_O}{[Ca]_C p_C} + 1\right)} \quad (19)$$

Compositional data are derived from APT spectra extracted from each material within the reconstruction, >1 nm from the 50% Ca isoconcentration surface (Table 5). These spectra integrate over the entire material and do not take into account the spatial variability evident in the proxygram. However, the increase in counting statistics inherent in these whole-region spectra offer higher confidence in the calculated Mg and Na content of the two materials, compared to the proxygram data.

The $\frac{[Ca]_O}{[Ca]_C}$ is calculated using volume-normalized counts, to account for the different volumes of materials captured in the APT volume. The R values are calculated from the raw counts of Na, Mg and Ca from each material, because the count ratio is independent of material volume. The parameters used in equation 19 can be seen in Table 6.

Table 5: Calcite and organic compositions measured by APT.

Element	Calcite (22797 nm ³)		Organic (17097 nm ³)	
	Counts	Counts nm ⁻³	Counts	Counts nm ⁻³
Ca	125971.1 ± 43.8	5.52578 ± 0.00192	1544.0 ± 17.7	0.09031 ± 0.00104
Mg	557.0 ± 95.5	0.02443 ± 0.00419	517.7 ± 54.1	0.03028 ± 0.00316
Na	940.3 ± 16.5	0.04125 ± 0.00072	405.4 ± 4.9	0.02371 ± 0.00029

Table 6: Parameters used in model (Eqn 19).

Parameter	Value
$\frac{[Ca]_O}{[Ca]_C}$	0.01634 ± 0.00019
$Mg R_C$	0.00442 ± 0.00076
$Mg R_O$	0.33530 ± 0.03525
$Na R_C$	0.00746 ± 0.00013
$Na R_O$	0.26256 ± 0.00437

4.4 Assumptions

This model relies on two key assumptions:

1. **The ionization efficiencies of Ca, Mg and Na are similar between calcite and organic materials in APT.** This is likely given the dependence of field evaporation on bond strength, and the tendency for these species to be ionically bound in both calcite and organic, which are both predominantly made of C, O or (in the case of organic) other low-Z elements.
2. **The entire organic layer has the same average composition as the ~10 nm layer measured by APT.** This untestable, but given the absence of any other data we have used this simple assumption.

4.5 Interpreting Model Output

The immediate output of the model is the counts M / count Ca signal associated with scanning the simulated beam across the simulated organic layer. This cannot be directly compared in terms of 'composition' because the APT data cannot be calibrated, or in terms of 'raw counts' because the ionization mechanisms of the techniques are fundamentally different. Furthermore, the efficiencies of Ca, Na and Mg are likely to be different in APT and ToF-SIMS. To overcome this, we consider the results of the simulation in terms of 'increase above background', where the organic-derived Mg/Ca or Na/Ca signal is normalized to the adjacent calcite region. This accounts for differences in the ionization efficiency between Ca, Na and Mg in APT, as long as the ionization efficiencies of each element remains relatively constant between the organic and calcite regions of the APT specimen.

A further consideration in interpreting the model output is the double Mg maxima either side of the POS position observed in our specimen (Fig 12A). These maxima are most likely

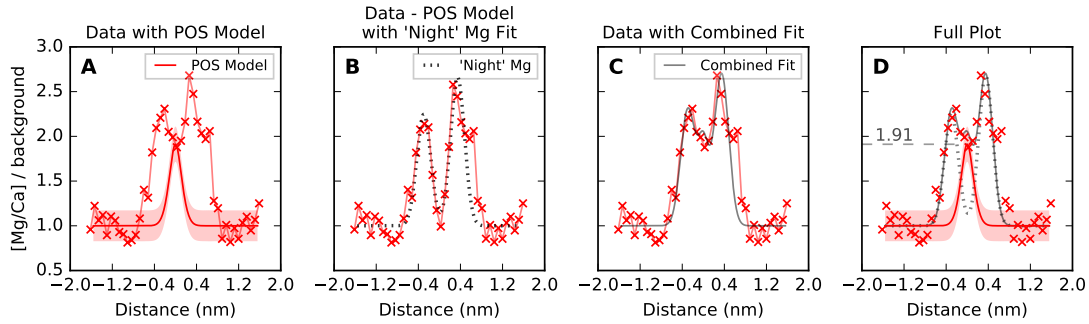


Figure 12: The components of the fit to the POS-associated Mg maximum. (A) Shows the ToF-SIMS profile data compared to the modelled POS signal, which under-estimates the data. (B) Shows the difference between the ToF-SIMS profile and the POS model, which is described well by two 312 nm Gaussian peaks. This is consistent with these peaks being caused by two thin layers of higher-Mg calcite approximately 300 nm away from the POS. The contribution of this Mg signal to the POS is negligible. (C) Shows the combination of the POS model and the adjacent Gaussians compared to the data. (D) Shows the complete fit, including both components.

associated with two layers of high-Mg ‘night’ calcite deposited on either side of the POS shortly after the start of calcification (~ 300 nm away from the POS). The distance between these peaks is ~ 2 times greater than the minimum resolvable distance of ToF-SIMS, so they can be treated as independent Mg enrichment signals, which are not associated with the POS. Because the distance of these peaks is ~ 1 beam FWHM away from the POS location, their contribution to POS counts should be negligible, and the POS signal should have little effect on these peaks. Therefore, these peaks may be included alongside the modelled POS signal to quantitatively explain the POS-associated Mg maxima in this specimen (Fig 12B). To achieve this, two Gaussians may be added to the modelled POS Mg signal, described by:

$$y = G(x, \text{amplitude}, \text{center}, \text{fwhm}) \quad (20)$$

$$= G(x, h, 0 - d, \text{fwhm}) + G(x, h \times dh, 0 + d \times dd, \text{fwhm}) \quad (21)$$

Where d and h are the distance from the POS and the peak height, and dd and dh are factors that allow the outer Mg maxima to be further from the POS and more intense, as expected based on the asymmetry of foraminiferal calcification. These parameters are fit to the data after subtraction of the modelled POS signal (Fig 12B). The double peak background was best fit by $d = 301.8 \pm 19.2$, $dd = 1.34 \pm 0.15$, $h = 1.26 \pm 0.12$, $dh = 1.16 \pm 0.09$. The multiplicative factors dd and dh are consistent with the magnitude of calcification asymmetry in *O. universa*.

Within this framework, we are able to reproduce both the POS-associated Mg/Ca and Na/Ca maxima observed by ToF-SIMS using a simulated POS width of 130 nm (Fig 12 C & D). Because our model is able to reproduce *both* signals with a single parameterization, we may accept our hypothesis. This tells us that the POS-associated signal observed in ToF-SIMS can be caused by a 130 nm organic layer of similar composition to the 10 nm organic layer observed by APT. Alongside the qualitative similarity between APT and ToF-SIMS results, the sampling location of the APT specimen, and in context of what is known about organic layers in foraminiferal calcite, this offers compelling evidence that the organic

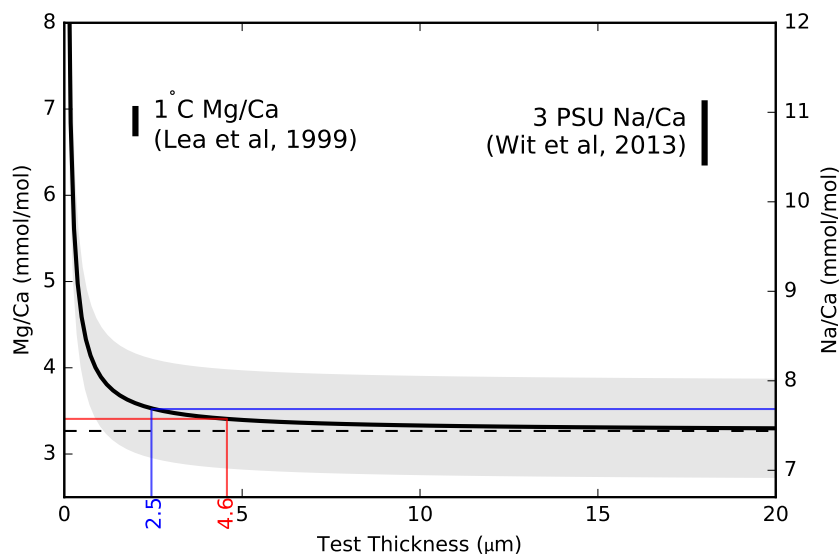


Figure 13: The effect of the POS on bulk foraminifera composition as a function of shell thickness. Red and blue lines indicate the shell thickness below which there is a greater than 0.5 C increase in Mg/Ca derived temperature, or a greater than 1 PSU increase in Na/Ca derived salinity, respectively.

measured by the APT is the POS, and that the POS is driving the simultaneous Na/Ca and Mg/Ca enrichment observed in ToF-SIMS.

Our hypothesis is tested by the ability of our model to simultaneously reproduce both the Na/Ca and Mg/Ca enrichments from a single parameterization. The POS width used in the parameterization does not influence the test of the hypothesis, and offers independent information on the amount of organic material required to cause the observed ToF-SIMS signal. If the entire organic layer has the same average composition as the organic layer measured by APT, this is consistent with a POS width of ~ 130 nm. This is in line with previous estimates of POS width (~ 100 nm).

4.6 Extension to Bulk Geochemistry

The end member mixing model (Section 4.3.2) can be extended to explore the contribution of organic layers to the overall skeletal chemistry of *O. universa*. This is achieved by using a simple geometric model of *O. universa* calcification to determine the p_O and p_C terms in Equation 18.

The *O. universa* chamber is described as two concentric spheres, a set distance (t) apart, containing a POS of known thickness (t_{POS}). At the start of calcification $t = t_{POS} = 130$ nm, and as calcite is added to the sphere $t > t_{POS}$. Thus, the volume of the POS is

constant, while the volume of calcite increases cubically with shell thickness:

$$S(r) = \frac{4}{3}\pi r^3 \quad (22)$$

$$V_{POS} = S(r_0 + t_{POS}/2) - S(r_0 - t_{POS}/2) \quad (23)$$

$$r_i = r_0 - t_{POS}/2 - \frac{t}{1+f} \quad (24)$$

$$r_o = r_0 + t_{POS}/2 + \frac{ft}{1+f} \quad (25)$$

$$V_{CAL} = S(r_o) - S(r_i) - V_{POS} \quad (26)$$

Where the volume (V) of each material is determined by the subtraction of two spheres whose radii (r) differ by a set distance. For the POS, the difference between the radii is determined by POS thickness (t_{POS}), which is evenly distributed either side of the initial radius (r_0). For calcite, the radii change with shell thickness (t), with a greater change on the outside of the shell (r_o) than the inside (r_i) to simulate the asymmetry of calcification. The asymmetry is described by a factor (f) that is the rate of external thickening divided by the rate of internal thickening. Thus, p_O and p_C are then calculated as:

$$p_O = \frac{V_{POS}}{V_{POS} + V_{CAL}} \quad (27)$$

$$p_C = \frac{V_{CAL}}{V_{POS} + V_{CAL}} \quad (28)$$

Using typical size and geometry parameters for an *O. universa* sphere ($r_0 = 200 \mu\text{m}$, $f = 3.0$) and the POS width estimated from the ToF-SIMS scale-up model ($t_{POS} = 130 \text{ nm}$), this can be used to predict the contribution of the POS to bulk foraminiferal composition, as a function of shell thickness (Fig 13).

4.7 Convolution Calculation

The p_O and p_C terms (Eqn 18) for the simulated ToF-SIMS analysis beam are calculated numerically in python.

```

1 # python 3.5
2 import numpy as np
3
4 def gauss(x, *p):
5     """
6     Gaussian function.
7     Parameters:
8     x: array-like
9     *p: parameters unpacked to A, mu, sigma
10    A: area
11    mu: centre
12    sigma: width
13
14    Note:
15    FWHM = 2 * sigma * np.sqrt(2 * np.log(2))
16    """
17    A, mu, sigma = p
18    return A * np.exp(-(x - mu)**2 / (2 * sigma**2))
19
20 # define parameters
21 o_width = 130 # width of organic layer
22 beam_fwhm = 312 # width of gaussian

```

```
23 window = 3 # width (n * fwhm) of the convolution
24
25 # generate distance (x) scale for convolution
26 ds = np.linspace(-n * beam.fwhm, n * beam.fwhm, 2 * n * beam.fwhm)
27
28 # calculate sigma for input into gaussian function
29 beam_sigma = beam.fwhm / (2 * np.sqrt(2 * np.log(2)))
30
31 # generate gaussian beam function
32 beam = gauss(ds, 1, 0, beam_sigma)
33 # normalise gaussian so integral = 1
34 nbeam = beam / beam.sum()
35
36 # generate organic layer function (tophat)
37 org = np.zeros(o_width * 3)
38 org[o_width : 2 * o_width] = 1.
39
40 # convolve beam with organic layer
41 c = np.convolve(nbeam, org, mode='same')
42
43 # The convolution of the gaussian whose integral is equal to 1
44 # and the tophat function ('c') is the proportion of the total
45 # illuminated area that contains organic material as the beam
46 # scans across the organic layer.
```

5 Data Processing Code

5.1 Data

Data Availability All data required to reproduce our analysis and raw APT data files are available as an additional supplement to the main manuscript.

Data Format Data files are in .csv format. Proxygram data contain distance and ion count columns, mass spectral data contain Da (mass/charge ratio) and count data.

5.2 Python Environment

All analyses are performed using the Jupyter Notebook environment. Non-base python packages are:

```
ipython==4.2.0
lmfit==0.9.3
matplotlib==1.5.1
numpy==1.10.2
pandas==0.18.1
scipy==0.16.1
uncertainties==2.4.6.1
```

To initialise the analysis environment, import the following:

```
1 import re
2 import time
3 import numpy as np
4 import pandas as pd
5 pd.options.mode.chained_assignment = None
6 import uncertainties.unumpy as un
7 import matplotlib.pyplot as plt
8
9 from scipy.optimize import curve_fit
10
11 %matplotlib inline
12
13 # For readability, helper functions for these analyses
14 # are stored in 'APT-helper-fns.py', which should be
15 # placed in the working directory
16 import APT_helper_fns as apt
17
18 # uncertainties helper functions
19 def nom(a):
20     try:
21         return un.nominal_values(a)
22     except:
23         return a
24
25 def err(a):
26     try:
27         return un.std_devs(a)
28     except:
29         return a
```

5.3 Compositional Analysis

5.3.1 Load Spectra and Peak Ranges

```

1 # load mass spectra from calcite and organic
2 # calcite volume
3 mscal = pd.read_csv('data/R31.04588-v02-calcite-v22797.csv')
4 mscal.columns = ['da', 'uc', 'c', 'bkg']
5 # organic volume
6 msorg = pd.read_csv('data/R31.04588-v02-organic-v17097.csv')
7 msorg.columns = ['da', 'uc', 'c', 'bkg']
8
9 # volume of two regions (nm3), noted in end of file name
10 calvol = 22797.
11 orgvol = 17097.
12
13 # load range file
14 ranges = apt.read_rrng('data/interface.bkg_corr_checked.RRNG')
15
16 # isolate pre-, pos- and peak- windows from range file
17 pre = ranges.iloc[["pre" in i for i in ranges.name.tolist()],:]
18 post = ranges.iloc[["post" in i for i in ranges.name.tolist()],:]
19 pk = ranges.iloc[["_" in i for i in ranges.name.tolist()],:]
20
21 pk.index = np.arange(len(pre))
22 pre.index = np.arange(len(pre))
23 post.index = np.arange(len(pre))
24
25 post.loc[:, 'n'] = post.index.values
26 pre.loc[:, 'n'] = pre.index.values
27 pk.loc[:, 'n'] = pk.index.values
28
29 # combine 'pre' and 'post' regions into single backgrounds file
30 bkg = pd.merge(pre, post, on='n', suffixes=('_pre', '_post'))

```

5.3.2 Identify peak and background counts

Counts within the peak and background windows are summed, and the background error is determined.

Background error is based on the standard deviation of the background region, and summed alongside the background region:

$$err_{bkg} = \sqrt{\sum err_i^2 + err_{i+1}^2 \dots err_n^2} \quad (29)$$

where n is the number of background points. The background, and its associated error, is then corrected for differences in the number of points captured by each region, and subtracted from the summed peak counts, along with the associated error:

$$C_{corr} = C_{pk} - \frac{n_{pk}}{n_{bg}} C_{bg} \quad (30)$$

where c denotes total counts, and n denotes the number of histogram bins captured by the peak window.

```

1 # Calcite calculations
2 pk.loc[:, 'cal_pk_n'] = pk.apply(lambda x: apt.npoints(x, mscal), axis=1)
3 pk.loc[:, 'cal_bkg_n'] = (pre.apply(lambda x: apt.npoints(x, mscal), axis=1) +
4 post.apply(lambda x: apt.npoints(x, mscal), axis=1))
5
6 pk.loc[:, 'cal_uc'] = pk.apply(lambda x: apt.sig_sum(x, mscal), axis=1)
7 pk.loc[:, 'cal_bkg'] = bkg.apply(lambda x: apt.bkg_sum(x, mscal), axis=1)
8 pk.loc[:, 'cal_bkg_std'] = bkg.apply(lambda x: apt.bkg_std(x, mscal), axis=1)
9
10 # package the standard deviations
11 pk.loc[:, 'cal_ubkg'] = un.uarray(pk.loc[:, 'cal_bkg'], pk.loc[:, 'cal_bkg_std'])
12 # calculate background multiplication factor to account for uneven numbers of measured points
13 pk.loc[:, 'cal_bkg_f'] = pk.loc[:, 'cal_pk_n'] / pk.loc[:, 'cal_bkg_n']
14 # background subtraction
15 pk.loc[:, 'cal_c'] = pk.loc[:, 'cal_uc'] - pk.loc[:, 'cal_ubkg'] * pk.loc[:, 'cal_bkg_f']
16
17 # Organic calculations
18 pk.loc[:, 'org_pk_n'] = pk.apply(lambda x: apt.npoints(x, msorg), axis=1)
19 pk.loc[:, 'org_bkg_n'] = (pre.apply(lambda x: apt.npoints(x, msorg), axis=1) +
20 post.apply(lambda x: apt.npoints(x, msorg), axis=1))

```

```

21 pk.loc[:, 'org_uc'] = pk.apply(lambda x: apt.sig_sum(x, msorg), axis=1)
22 pk.loc[:, 'org_bkg'] = bkgs.apply(lambda x: apt.bkg_sum(x, msorg), axis=1)
23 pk.loc[:, 'org_bkg_std'] = bkgs.apply(lambda x: apt.bkg_std(x, msorg), axis=1)
24
25 # package the standard deviations
26 pk.loc[:, 'org_ubkg'] = un.uarray(pk.loc[:, 'org_bkg'], pk.loc[:, 'org_bkg_std'])
27 # calculate background multiplication factor to account for uneven numbers of measured points
28 pk.loc[:, 'org_bkg_f'] = pk.loc[:, 'org_pk_n'] / pk.loc[:, 'org_bkg_n']
29 # background subtraction
30 pk.loc[:, 'org_c'] = pk.loc[:, 'org_uc'] - pk.loc[:, 'org_ubkg'] * pk.loc[:, 'org_bkg_f']

```

5.3.3 Na and Mg Interference Corrections

Na is calculated by subtracting predicted $^{46}\text{Ca}^{2+}$ counts from the 23 Da peak, based on $^{40}\text{Ca}^{2+}$ counts.

Mg is calculated by two approaches:

1. Subtraction of predicted $^{46}\text{Ca}^{2+}$ counts from the 24 Da peak
2. Prediction of total Mg from the minor ^{25}Mg and ^{26}Mg peaks

The following outlines the evaluation of these different Mg selection methods, and the selection of the Mg correction method outlined in Figure 7.

Note: The natural abundances of Mg isotopes are .79, .10 and .11 for ^{24}Mg , ^{25}Mg and ^{26}Mg , respectively. Total Mg counts may therefore be calculated by either:

$${}^T\text{Mg}_c = {}^{24}\text{Mg}_c \frac{1}{.79} = {}^{25}\text{Mg}_c \frac{1}{.10} = {}^{26}\text{Mg}_c \frac{1}{.11} \quad (31)$$

Ca subtraction from Mg^+ and Na^+ peaks. The counts of less abundant Ca isotopes are predicted from $^{40}\text{Ca}^{2+}$.

```

1 # Specify Ca isotope abundances:
2 caiso = {'40Ca': 96.941000000000003,
3         '42Ca': 0.64700000000000002,
4         '43Ca': 0.13500000000000001,
5         '44Ca': 2.08599999999999999,
6         '46Ca': 0.0040000000000000001,
7         '48Ca': 0.187}
8 # isolate Ca 2+ peaks
9 opks = pk.loc[(pk.centre >= 19.5) & (pk.centre <= 24.5),
10             ['centre', 'cal_c', 'org_c']].copy()
11
12 opks

```

	centre	cal_c	org_c
15	19.9657	76146.1+/-4.7	680.7+/-2.9
16	20.9783	531.0+/-21.2	20.0+/-2.7
17	21.9745	1600.0+/-17.7	6.0+/-4.0
18	22.9785	943.4+/-16.5	405.5+/-4.9
19	23.9763	214.0+/-10.2	197.0+/-3.9

```

1 # predict Ca peak intensity from 40Ca counts
2 opks.loc[:, 'ca_abund'] = [caiso[p] for p in ['40Ca', '42Ca', '44Ca', '46Ca', '48Ca']]
3 opks.loc[:, 'cal_pred'] = (opks.loc[:, 'ca_abund'] *
4                          opks.loc[15, 'cal_c'] /
5                          opks.loc[:, 'ca_abund'].max())
6 opks.loc[:, 'org_pred'] = (opks.loc[:, 'ca_abund'] *
7                          opks.loc[15, 'org_c'] /
8                          opks.loc[:, 'ca_abund'].max())
9
10 # plot code omitted

```

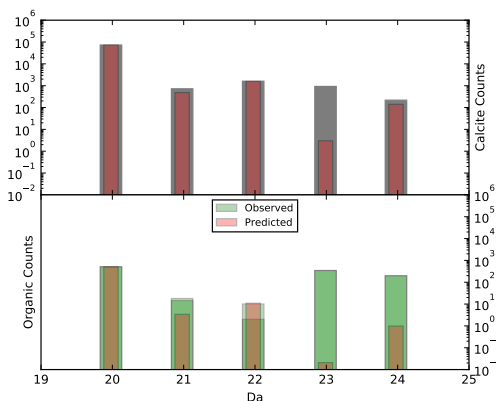


Figure 14: The measured (green/grey) counts for ToF-peaks between 19.5-24.5 Da from organic and calcite spectra, compared to the predicted counts (red) based on the $^{40}\text{Ca}^{2+}$ peak. For measured counts, the difference between the two bars (e.g. at 22 Da in organic) represents the standard deviation of the counts.

```

1 # subtract predicted from observed counts
2 opks.loc[:, 'org-sub'] = opks.loc[:, 'org-c'] - opks.loc[:, 'org-pred']
3 opks.loc[:, 'cal-sub'] = opks.loc[:, 'cal-c'] - opks.loc[:, 'cal-pred']
4
5 # isolate Ca subtracted Na and Mg counts
6 cacorr = opks.loc[[18, 19], ['centre', 'cal-c', 'cal-pred', 'cal-sub', 'org-c', 'org-pred', 'org-sub']]
7
8 # correct for total Mg counts
9 cacorr.loc[19, 'cal-tMg'] = cacorr.loc[19, 'cal-sub'] * 1 / .79
10 cacorr.loc[19, 'org-tMg'] = cacorr.loc[19, 'org-sub'] * 1 / .79
11
12 print('Ca Subtracted\n' +
13       'Na counts\n   Calcite: {:.1f}\n   Organic: {:.1f}'.format(*cacorr.loc[18, ['cal-sub', 'org-sub']]))
14       '\n Mg counts\n   Calcite: {:.1f}\n   Organic: {:.1f}'.format(*cacorr.loc[19, ['cal-tMg', 'org-tMg']]))

```

Ca Subtracted

Na counts

Calcite: 940.3+/-16.5

Organic: 405.4+/-4.9

Mg counts

Calcite: 85.0+/-13.0

Organic: 247.7+/-5.0

Mg⁺ counts from $^{25}\text{Mg}^+$ and $^{24}\text{Mg}^+$ peaks. Peaks at 24, 25 and 26 Da are evident in both materials, corresponding to monovalent Mg ions. The peak at 24 suffers from a significant interference from $^{48}\text{Ca}^{2+}$, and cannot be directly used to measure $^{24}\text{Mg}^+$. There may be polyatomic interferences at 25 and 26 Da, but these are impossible to constrain. A conservative estimate of Mg counts is therefore obtained from the lesser of the 25 and 26 peaks.

```

1 # Specify Mg isotope abundances:
2 mgiso = {'24Mg': 0.7899,
3          '25Mg': 0.1000,
4          '26Mg': 0.1101}
5
6 # isolate peaks of interest (24 & 25 Da)

```



```

7 mg1 = pk.loc[(pk.centre >= 24.5) & (pk.centre <= 26.5),
8           ['centre', 'cal_c', 'org_c']].copy()
9
10 # calculate total Mg counts based on their intensity
11 mg1.loc[:, 'org_tMg'] = mg1.loc[:, 'org_c'] * [1/mgiso['25Mg'], 1 / mgiso['26Mg']]
12 mg1.loc[:, 'cal_tMg'] = mg1.loc[:, 'cal_c'] * [1/mgiso['25Mg'], 1 / mgiso['26Mg']]
13
14 mg1.loc[:, ['centre', 'cal_c', 'cal_tMg', 'org_c', 'org_tMg']]
15
16 print('25Mg+ count predicted\n   Calcite: {:.1f}\n   Organic: {:.1f}'.format(*mg1.loc[45, ['cal_tMg', 'org_tMg']]))
17 print('\n26Mg+ count predicted\n   Calcite: {:.1f}\n   Organic: {:.1f}'.format(*mg1.loc[46, ['cal_tMg', 'org_tMg']]))

```

```

25Mg+ count predicted
   Calcite: 300.0+/-82.7
   Organic: 460.0+/-33.1
26Mg+ count predicted
   Calcite: 154.5+/-52.7
   Organic: 418.2+/-27.5

```

Mg²⁺ counts from ²⁵Mg²⁺ peak Peaks at 12 and 13 Da are clear in both materials, and calcite has a clear 12.5 Da peak. The 12 and 13 peaks suffer from interference by ¹²C⁺ and ¹²CH⁺, and cannot be reliably assigned to ²⁴Mg²⁺ or ²⁶Mg²⁺. Total Mg counts are therefore estimated using the 12.5 Da peak, which is attributable to ²⁵Mg²⁺.

```

1 # isolate 12.5 peak
2 mg2 = pk.loc[(pk.centre >= 12.3) & (pk.centre <= 12.7),
3           ['centre', 'cal_c', 'org_c']].copy()
4
5 # calculate total Mg counts
6 mg2.loc[:, 'cal_tMg'] = mg2.loc[:, 'cal_c'] * 1 / mgiso['25Mg']
7 mg2.loc[:, 'org_tMg'] = mg2.loc[:, 'org_c'] * 1 / mgiso['25Mg']
8
9 print('25Mg++ count predicted\n   Calcite: {:.1f}\n   Organic: {:.1f}'.format(*mg2.loc[44, ['cal_tMg', 'org_tMg']]))

```

```

25Mg++ count predicted
   Calcite: 257.0+/-47.7
   Organic: 57.7+/-42.7

```

Calculation of Mg Counts A calculated Mg count estimate must include both single and double charged ions. Double charged ions may only be estimated by prediction from the 12.5 peak. Singly charged ions may be estimated by Ca subtraction, or prediction from the 25 or 26 Da peak.

```

1 # compile all Mg estimates
2 mgpred = mg1.append(cacorr.loc[19, ['centre', 'cal_tMg', 'org_tMg']])
3
4 # bring in Mg2+
5 mgpred.loc[:, 'cal_tMg2'] = mg2.loc[:, 'cal_tMg'].values
6 mgpred.loc[:, 'org_tMg2'] = mg2.loc[:, 'org_tMg'].values
7
8 # calculate sum
9 mgpred.loc[:, 'cal_sMg'] = mgpred.loc[:, 'cal_tMg'] + mg2.loc[:, 'cal_tMg'].values
10 mgpred.loc[:, 'org_sMg'] = mgpred.loc[:, 'org_tMg'] + mg2.loc[:, 'org_tMg'].values
11
12 # plot code omitted.

```

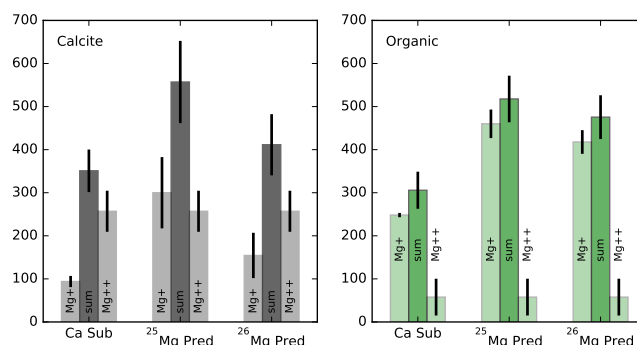


Figure 15: The total Mg counts estimated via three different methods. The central, darker bar is the sum Mg counts, bracketed by the Mg^+ and Mg^{2+} component peaks. All peaks contain the Mg^{2+} counts predicted from the 12.5 Da peak, and differ in their Mg^+ component: in 'Ca Sub' the single charge contribution is calculated by subtraction of the interfering $^{48}Ca^{2+}$ counts, and in ' ^{25}Mg Pred' and ' ^{26}Mg Pred' they are extrapolated from the $^{25}Mg^+$ or $^{26}Mg^+$ peaks, respectively.

```

1 print('Ca Sub\n Calcite: {:.1f}\n Organic: {:.1f}\n'.format(*mgpred.loc[19,['cal_sMg', 'org_sMg']]) +
2       '^25Mg Pred\n Calcite: {:.1f}\n Organic: {:.1f}\n'.format(*mgpred.loc[45,['cal_sMg', 'org_sMg']]) +
3       '^26Mg Pred\n Calcite: {:.1f}\n Organic: {:.1f}\n'.format(*mgpred.loc[46,['cal_sMg', 'org_sMg']]))

```

Ca Sub

Calcite: 342.0+/-49.4
Organic: 305.4+/-43.0

^{25}Mg Pred

Calcite: 557.0+/-95.5
Organic: 517.7+/-54.1

^{26}Mg Pred

Calcite: 411.6+/-71.0
Organic: 475.8+/-50.8

Observations:

- *Ca Sub*: In both materials, the Ca subtracted $^{24}Mg^+$ peak gives a lower total Mg count than the other estimates of Mg^+ . The propagated error on the Ca subtraction is relatively small, but this does not reflect uncertainties in the shape of peaks - e.g. the $^{40}Ca^{2+}$ peak has a large tail, compared to the other Ca^{2+} peaks. This could introduce large uncertainties in this background correction, particularly where the interference peak is a significant portion of the observed peak (i.e. this matters more for $^{48}Ca^{2+}$, than $^{46}Ca^{2+}$; Fig 14).
- *Mg Pred* peaks: In calcite, ^{25}Mg Pred and *Ca Pred* estimates are within error of each other. ^{25}Mg Pred is larger, possibly suggesting a minor polyatomic interference at 26 Da? In organic, ^{25}Mg Pred and ^{26}Mg Pred are similar, and both larger than the *Ca sub* estimate. The same Mg^{2+} count estimate is used, so differences are caused by variation in the Mg^+ estimate.

Choice of Mg Correction. The ^{26}Mg *Pred* is used to estimate Mg counts throughout because:

1. The *Ca sub* method deals with the subtraction of peaks with complex, variable shapes, and non-equivalent selection regions of the peaks may bias the subtraction in unconstrained ways. Because ^{48}Ca makes a significant contribution to the 24 Da peak, this may be important here (as opposed to for Na, where the peak is almost entirely Na counts).
2. In both calcite and organic the 26 Da peak yields the lowest counts of the two minor Mg peaks, and is the most conservative.
3. In both calcite and organic, the ^{26}Mg *Pred* value is within error of another estimate (^{25}Mg *Pred* in organic, *Ca Pred* in calcite). This gives confidence that it is representative of each material, while allowing the use of the same Mg correction method in both materials.

5.3.4 Apply Mg and Na Corrections

```

1 # Na counts
2 na = pd.DataFrame({'name': 'R58.Na1',
3                   'cal_c': cacorr.loc[18, 'cal_c'],
4                   'org_c': cacorr.loc[18, 'org_c'],
5                   'centre': 23.},
6                   index=[18.1])
7
8 # Mg counts
9 mg = pd.DataFrame({'name': 'R61.Mg1',
10                  'cal_c': mgpred.loc[46, 'cal_sMg'],
11                  'org_c': mgpred.loc[46, 'org_sMg'],
12                  'centre': 24.},
13                  index=[19.1])
14
15 # Add these to the count dataframe
16 pk = pd.concat([pk, na, mg])
17
18 # remove the Na and Mg counts from the minor Ca peaks
19 pk.loc[[18, 19], 'cal_c'] = cacorr.loc[[18,19], 'cal_c']
20 pk.loc[[18, 19], 'org_c'] = cacorr.loc[[18,19], 'org_c']

```

5.3.5 Calculate Material Compositions

First, complex ions must be decomposed into their constituent elements so that they may be counted correctly.

```

1 # Find which elements are present
2 nmisc = [re.sub('R[0-9]+_?[Name]p[0-9]+', '', n) for n in pk.name]
3 ccomps = [apt.composition(n) for n in nmisc]
4 elcont = apt.composition('').join(nmisc).keys()
5
6 # sum the counts of each element in each material.
7 # 1. create an empty dataframe for counts of all these elements
8 comp = pd.DataFrame(0, index=elcont, columns=['cal', 'org'])
9
10 # 2. populate the dataframe with counts for each element
11 for e in elcont:
12     M = []
13     for c in ccomps:
14         if e in c.keys():
15             M.append(c[e])
16         else:
17             M.append(0)
18     for n in np.arange(len(M)):
19         comp.loc[e, 'cal'] += pk.cal_c.iloc[n] * M[n]
20         comp.loc[e, 'org'] += pk.org_c.iloc[n] * M[n]
21

```

```

22 # 3. calculate total counts
23 comp['total'] = comp.loc[:, 'cal'] + comp.loc[:, 'org']
24
25 # 4. calculate percentage of total counts
26 cpc = comp.copy()
27 cpc.loc[:, 'org'] = cpc.loc[:, 'org'].values / np.sum(cpc.loc[:, 'org'].values)
28 cpc.loc[:, 'cal'] = cpc.loc[:, 'cal'].values / np.sum(cpc.loc[:, 'cal'].values)
29 cpc.loc[:, 'total'] = cpc.loc[:, 'total'].values / np.sum(cpc.loc[:, 'total'].values)
30
31 cpc = cpc * 100
32
33 # 5. calculate M/Ca ratios
34 rats = comp / comp.loc['Ca', :]
35 rats.drop('Ca', inplace=True)
36
37 # Na/Ca and Mg/Ca ratios in cts/kcts Ca
38 rats.loc[['Mg', 'Na'], :] * 1E3

```

	cal	org	total
Mg	4.4+/-0.7	313.0+/-28.9	9.5+/-0.8
Na	7.43+/-0.13	189.0+/-2.4	10.45+/-0.13

5.4 Data Quality Checks

5.4.1 On-Peak Counts

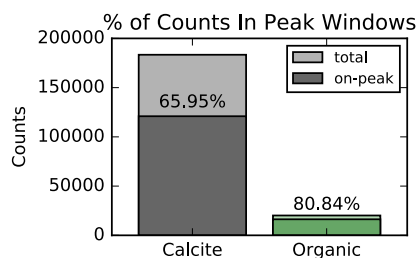


Figure 16: This offers an estimate of the rough 'signal/background' ratio for the materials. A lower proportion of counts in calcite are 'on-peak', because of the large $^{40}\text{Ca}^{2+}$ tail, which gives a high background.

5.4.2 Expected vs. Observed Efficiency

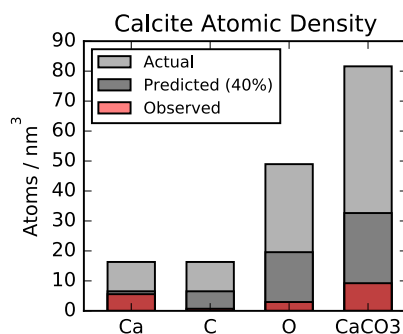


Figure 17: This plot shows the actual atoms/nm³ for Ca, C and O in calcite, the predicted density based on the nominal 40% efficiency of the LEAP 4000, and the measured atomic density of calcite by APT. Ca recovery is close to the nominal instrument efficiency, while C and O recovery is much lower. Note that the ratio of O:C is ~ 3: C and O loss is in line with CO₃ stoichiometry, perhaps implying the mid-flight decomposition of CO₃ groups into neutral products?

5.5 Proxygram Processing

```

1 # import raw proxygram data: raw count data from each of the ranges defined in the *.rimg* file above, for
  0.01nm bins perpendicular to the organic/mineral interface (50% Ca isosurface).
2 pg = pd.read_csv('data/R31_04588-v02-artefactGone.interface40Ca32_full-bkg-proxygram-Mg25_12nm_0.01
  bin_ioncounts.csv')
3
4 # separate pre, post and pk proxygrams
5 nms = pg.columns.tolist()[1:-1]
6 cpre = ['Distance (nm)'] + [n for n in nms if 'pre' in n]
7 cpost = ['Distance (nm)'] + [n for n in nms if 'post' in n]
8 cpk = ['Distance (nm)'] + [n for n in nms if 'post' not in n and 'pre' not in n]
9
10 pgpre = pg[cpre]
11 pgpost = pg[cpk]
12 pgpk = pg[cpk]
13
14 # rename columns so they match
15 pgpre.columns = [pgpre.columns[0]] + [n.replace('pre','') for n in pgpre.columns[1:].tolist()]
16 pgpost.columns = [pgpost.columns[0]] + [n.replace('post','') for n in pgpost.columns[1:].tolist()]
17
18 # combine background counts
19 pgbkg = pgpre + pgpost
20
21 # make column names the same as peaks
22 o = pgbkg.columns.tolist()
23 n = pgpk.columns.tolist()
24 nc = ['Distance (nm)']
25 for oi in o:
26     for ni in n:
27         if oi + "-" in ni:
28             nc.append(ni)
29 pgbkg.columns = nc
30
31 # apply background correction
32 pgc = pgpk.subtract(pgbkg, axis=1)
33 # replace distance scale
34 pgc['Distance (nm)'] = pg['Distance (nm)']
35 # get total count and dstep data back
36 pgc['Sample Count'] = pg['Sample Count']
37 # pgc['dstep'] = pg['dstep']

```

5.5.1 Na Peak Overlap Correction

Corrected by subtracting predicted $^{46}\text{Ca}^{2+}$ counts from the 23 Da peak, as above.

```

1 # isolate 20 and 23 Da proxygrams
2 nac = pgc.loc[:, ['Distance (nm)', 'R46-Ca1', 'R58-Ca1']].copy()
3 # predict Ca46 counts
4 nac.loc[:, 'Ca-pred'] = nac.R46-Ca1 * caiso['46Ca'] / caiso['40Ca']
5 # calculate Na counts
6 nac.loc[:, 'p1Na1'] = nac.loc[:, 'R58-Ca1'] - nac.loc[:, 'Ca-pred']
7 # calculate remaining Ca counts
8 nac.loc[:, 'R58-Ca1'] = nac.loc[:, 'Ca-pred']
9
10 # transfer predicted Na counts to pgc
11 pgc.loc[:, 'p1Na1'] = np.nan # create empty column
12 pgc.loc[:, ['R58-Ca1', 'p1Na1']] = nac.loc[:, ['R58-Ca1', 'p1Na1']]

```

5.5.2 Calculation of Mg Proxygram

Peak Choice. Because the 12.5 Da peak is not obvious in the mass spectrum of the total organic region (Fig 7), it is likely that this peak will be undetectable in a proxygram, where the total number of ions in each bin is orders of magnitude less than the entire region. Therefore, we investigate the proxygrams for the 'peak' and 'background' regions independently, to evaluate whether or not to include the 12.5 Da peak in proxygram Mg calculation.

```

1 bkgcheck = apt.rebin(pg, 800, 'Sample Count')
2
3 # plot code omitted

```

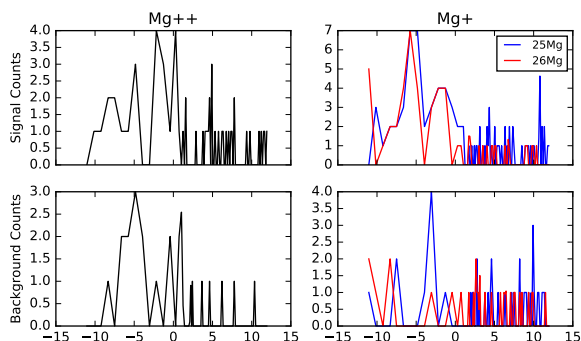


Figure 18: Large regions of the Mg⁺⁺ proxygram have a greater background than signal. Using the 12.5 Da peak to calculate Mg counts in the proxygram simply adds noise. The ²⁵Mg⁺ and ²⁶Mg⁺ signal counts both show the same pattern within the organic layer, giving confidence that they do represent Mg counts. The ²⁶Mg⁺ has the lower background, and will be used in the proxygram

Only the ²⁶Mg⁺ peak will be used for generating the Mg proxygram.

5.5.3 Mg Extrapolation from ²⁶Mg⁺

```

1 # find names of peaks
2 nms = pk.loc[(pk.centre < 26.5) & (pk.centre > 25.5), 'name'].values
3 nms = np.concatenate([[ 'Distance (nm)' ], nms])
4
5 # isolate 26+ peak
6 mgc = pgc.loc[:, nms]
7
8 # calculate Mg+ counts
9 mgc.loc[:, 'mg1'] = mgc.loc[:, 'R142.Mg26'] * 1 / mgiso['26Mg']
10
11 # remove old Mg peaks labelled as 'Mg' in pgc
12 pgc = pgc.loc[:, [c for c in pgc.columns if 'Mg' not in c]]
13
14 # propagate changes to main proxygram dataframe
15 pgc.loc[:, 'p2Mg1'] = mgc.loc[:, 'mg1']

```

5.5.4 Decompose Complex Ions

```

1 # Find which elements are present
2 nms = [nm for nm in pgc.columns.tolist() if nm not in ['Distance (nm)', 'Sample Count', 'dstep']]
3 nmcs = [re.sub('R[0-9]+-?[Name][p[0-9]+', '', n) for n in nms]
4 ccomps = [apt.composition(n) for n in nmcs]
5 elcont = apt.composition('').join(nmcs).keys()
6
7 # Find total counts of all elements
8 pgc['total_atoms'] = 0
9
10 for e in elcont:
11     M = []
12     for c in ccomps:
13         if e in c.keys():
14             M.append(c[e])
15         else:
16             M.append(0)
17 # Multiply ion count by stoichiometry to get total
18 pgc.loc[:, 't'+e] = 0
19 for n in np.arange(len(M)):
20     pgc.loc[:, 't'+e] += pgc.loc[:, nms[n]].values * M[n]
21 pgc['total_atoms'] += pgc['t'+e]

```

5.5.5 Atomic Yield Differences

Calcite yields 10x more atoms than the organic material (Fig 19).


```

39 com['n'] = tmp[1]
40 # calculate the atomic density
41 com['mol'] = com.density/com.mw
42 com['at'] = com.mol * avo * com.n
43 com['nm3'] = com['at'] / 1E21
44
45 # repeat for the Robbins & Brew compounds
46 tmp = list(zip(*robbins.formula.map(MWcalc)))
47 robbins['mw'] = tmp[0]
48 robbins['n'] = tmp[1]
49 # calculate atomic density
50 robbins['mol'] = robbins.density/robbins.mw
51 robbins['at'] = robbins.mol * avo * robbins.n
52 robbins['nm3'] = robbins['at'] / 1E21
53
54 # plot code omitted

```

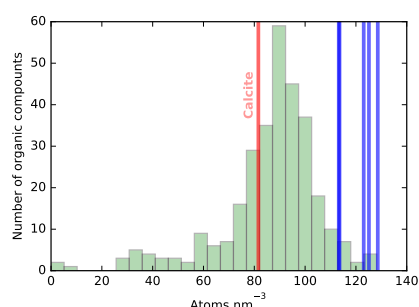


Figure 20: The atomic density of a range of organic compounds (green histogram), amino acids known to be associated with foraminiferal calcite (blue lines) and calcite (red line).

The majority of organic molecules are **more** dense than calcite, particularly the amino acids identified as abundant in foraminiferal organics (Fig 20). Note, however, that these organic density estimates will be derived from crystallized or dried organics, and therefore may not reflect the *in-vivo* density of the material. Based on these data, we would expect an approximately similar atomic density from the organic and calcite materials - not 10x fewer from the organics, as we observe. **Therefore, we will normalise the data to total atom count, to normalise for differences in evaporation field between the materials.**

5.5.6 Rebin Proxygrams

The proxygrams are rebinned to maintain constant total atom count (i.e. normalise for atomic yield differences).

To perform this normalisation, and to produce statistically useful data, it is further useful to re-bin the proxygram data, such that each bin contains the same number of total atoms. This effectively normalises for differences in atomic yield, and provides similar count statistics for data on either side of the interface, allowing their direct comparison, at the cost of spatial resolution.

```

1 # isolate data of interest
2 pgs = pgc[['Distance (nm)'] + ['t' + e for e in elcont] + ['total_atoms']]
3
4 pga = apt.rebin(pgs, 800, 'total_atoms') # Note: rebinning function is in APT.helper.fns.py
5
6 # plot code omitted

```

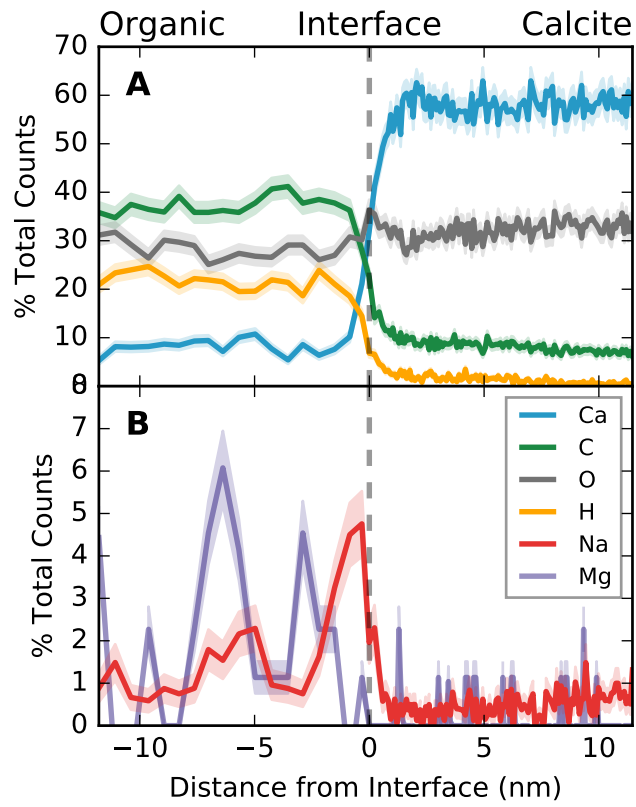



Figure 21: The rebinned proxygrams.

5.6 APT Helper Functions

The following contains a number of functions used numerous times in the above code. If you wish to execute any of the above, copy the functions below into a text file, name it 'APT_helper_fns.py', and place it in your python working directory.

```

1 # helper functions for APT analysis
2
3 import numpy as np
4 import pandas as pd
5 import re
6
7
8 def read_rrng(f):
9     """
10    Loads an IVAS .rrng file.
11
12    Returns
13    -----
14    ranges : pandas.DataFrame
15    """
16    import re
17
18    rf = open(f, 'r').read()
19
20    # extract relevant field using a regex find
21    ranges = re.findall('(Range[0-9]+)=+(\d+.\d+) +(\d+.\d+) +Vol:(\d+.\d+) +Name:(R[0-9]+.*) +Color:([A-Z0-9]{6})', rf)
22    ranges = pd.DataFrame(ranges, columns=['range', 'rmin', 'rmax', 'vol', 'name', 'color'])
23    ranges.iloc[:, 1:4] = ranges.iloc[:, 1:4].astype(float)
24    # define range centres and width
25    ranges['centre'] = (ranges.rmin + ranges.rmax) / 2
26    ranges['width'] = ranges.rmax - ranges.rmin
27
28    return ranges
29
30
31 # Functions for getting per-range counts from spectra.
32 def npoints(pk, df):
33     # return number of points measured
34     return len(df.loc[(df.da >= pk.rmin) & (df.da <= pk.rmax)])
35
36
37 def sig_sum(pk, df):
38     # return sum of points in single interval
39     return np.nansum(df.loc[(df.da >= pk.rmin) & (df.da <= pk.rmax), 'c'])
40
41
42 def bkg_sum(pk, df):
43     # return sum of points in 'pre' and 'post' background regions
44     pre = df.loc[(df.da >= pk.rmin_pre) & (df.da <= pk.rmax_pre)]
45     post = df.loc[(df.da >= pk.rmin_post) & (df.da <= pk.rmax_post)]
46     return np.nansum(pre.c) + np.nansum(post.c)
47
48
49 def bkg_std(pk, df):
50     # return standard deviation of background sum.
51     sub = df.loc[(df.da >= pk.rmin_pre) & (df.da <= pk.rmax_pre) |
52                (df.da >= pk.rmin_post) & (df.da <= pk.rmax_post)]
53     std = np.std(sub.c)
54     n = len(sub)
55     return np.sqrt(np.sum([std**2] * n))
56
57
58 # function to identify elemental contents from range name
59 def composition(comp):
60     ions = {}
61     els = ['Th', 'Pb', 'TI', 'Hg', 'Au', 'Pt',
62           'Ir', 'Os', 'Re', 'Ta', 'Hf', 'Lu',
63           'Yb', 'Tm', 'Er', 'Ho', 'Dy', 'Tb',
64           'Gd', 'Eu', 'Sm', 'Nd', 'Pr', 'Ce',
65           'La', 'Ba', 'Cs', 'Xe', 'Te', 'Sb',
66           'Sn', 'In', 'Cd', 'Ag', 'Pd', 'Rh',
67           'Ru', 'Mo', 'Nb', 'Zr', 'Sr', 'Rb',
68           'Kr', 'Br', 'Se', 'As', 'Ge', 'Ga',
69           'Zn', 'Cu', 'Ni', 'Co', 'Fe', 'Mn',
70           'Cr', 'Ti', 'Sc', 'Ca', 'Ar', 'Cl',
71           'Si', 'Al', 'Mg', 'Na', 'Ne', 'Be',
72           'Li', 'He', 'W', 'I', 'Y', 'V', 'K',
73           'S', 'P', 'F', 'O', 'N', 'C', 'B', 'H']
74     for e in els:

```

```

75     if e in comp:
76         ion = re.findall(e+"[0-9]?", comp)[0]
77         comp = comp.replace(ion, "")
78         if ion[-1].isalpha():
79             ion += "1"
80         ions[re.findall("[A-Za-z]+", ion)[0]] = \
81             int(re.findall("[0-9]+", ion)[0])
82     return ions
83
84
85 # Function to re-bin proxygram data
86 def rebin(pg, sc=800, axis='total_atoms', bins=None):
87     tmp = pg.cumsum() # cumulative sum of all counts along distance axis
88     tmp['Distance (nm)'] = pg['Distance (nm)'] # reapply original distance scale, to preserve space dimension
89     tmp.index = tmp[axis] # set index to cumulative sample count, for interpolation.
90     if bins is None:
91         bins = np.arange(0, tmp[axis].max(), sc) # if no specific bins are provided, create evenly spaced
92         # bins 'sc' apart.
93     new = tmp.reindex(bins) # make new index from 0 to max counts, at intervals of [sc]
94     new = pd.DataFrame(np.nan, index=bins, columns=tmp.columns) # create a blank dataframe with the new count
95     # intervals as the index.
96     new.loc[0] = 0
97     new.loc[0]['Distance (nm)'] = tmp['Distance (nm)'].iloc[0] + (tmp['Distance (nm)'].iloc[0] - tmp['
98     Distance (nm)'].iloc[1])
99
100    tmp = pd.concat([tmp,new]) # join the new index to the cumulative sum dataframe
101    tmp = tmp.sort_index() # sort the merged dataframe so values are ready to interpolate
102    tmp = tmp.interpolate(method='values') # interpolate the values onto the new count axis
103    tmp = tmp.loc[new.index] # only select the new ones
104    tmp = tmp.groupby(tmp.index).first().dropna() # get rid of any pesky double values (arise if any count %
105    cm == 0)
106    ds = tmp['Distance (nm)'] # record distance scale of interpolated values, before back-transformation
107
108    tmp = tmp.diff() # diff the data, to get back from cumulative sum to histogram space
109    tmp['dstep'] = tmp['Distance (nm)'] # record the diffed distance scale as a step size variable
110    tmp['Distance (nm)'] = ds # reinsert interpolated distance scale
111    tmp = tmp.iloc[1:] # chuck NaN values
112    return tmp

```

References

- O. Branson, S. A. T. Redfern, T. Tyliszczak, A. Y. Sadekov, G. Langer, K. Kimoto, and H. Elderfield. The coordination of Mg in foraminiferal calcite. *Earth Plan Sci Lett*, 383: 134–141, 2013.
- S. Eggins, P. de Deckker, and J. Marshall. Mg/Ca variation in planktonic foraminifera tests: implications for reconstructing palaeo-seawater temperature and habitat migration. *Earth Plan Sci Lett*, 212(3):291–306, 2003.
- S. Eggins, A. Y. Sadekov, and P. de Deckker. Modulation and daily banding of Mg/Ca in *Orbulina universa* tests by symbiont photosynthesis and respiration: a complication for seawater thermometry? *Earth Plan Sci Lett*, 225(3-4):411–419, 2004.
- R. I. Gabitov, A. C. Gagnon, Y. Guan, J. M. Eiler, and J. F. Adkins. Accurate Mg/Ca, Sr/Ca, and Ba/Ca ratio measurements in carbonates by SIMS and NanoSIMS and an assessment of heterogeneity in common calcium carbonate standards. *Chem Geol*, 356:94–108, 2013.
- L. M. Gordon and D. Joester. Nanoscale chemical tomography of buried organic–inorganic interfaces in the chiton tooth. *Nature*, 469(7329):194–197, 2011.
- B. T. Huber, J. Bijma, and H. J. Spero. Blue water SCUBA collection of planktic foraminifera. In M. A. Lang and C. C. Baldwin, editors, *Methods and Techniques of Underwater Research*. 1996.
- D. W. Lea and H. J. Spero. Experimental determination of barium uptake in shells of the planktonic foraminifera *Orbulina universa* at 22°C. *Geochim Cosmochim Acta*, 56(7): 2673–2680, 1992.
- F. Perez and B. E. Granger. IPython: A system for interactive scientific computing. *Computing in Science & Engineering*, 9(3):21–29, 2007.
- H. J. Spero. Ultrastructural examination of chamber morphogenesis and biomineralization in the planktonic foraminifer *Orbulina universa*. *Mar Biol*, 99(1):9–20, 1988.
- H. J. Spero, S. M. Eggins, A. D. Russell, L. Vetter, M. R. Kilburn, and B. Hönisch. Timing and mechanism for intratest Mg/Ca variability in a living planktic foraminifer. *Earth Plan Sci Lett*, 409(C):32–42, 2015.
- A. L. Stancik and E. B. Brauns. A simple asymmetric lineshape for fitting infrared absorption spectra. *Vib Spectrosc*, 47(1):66–69, 2008.