nature astronomy

Article

https://doi.org/10.1038/s41550-023-02160-0

Detection of HCN and diverse redox chemistry in the plume of Enceladus

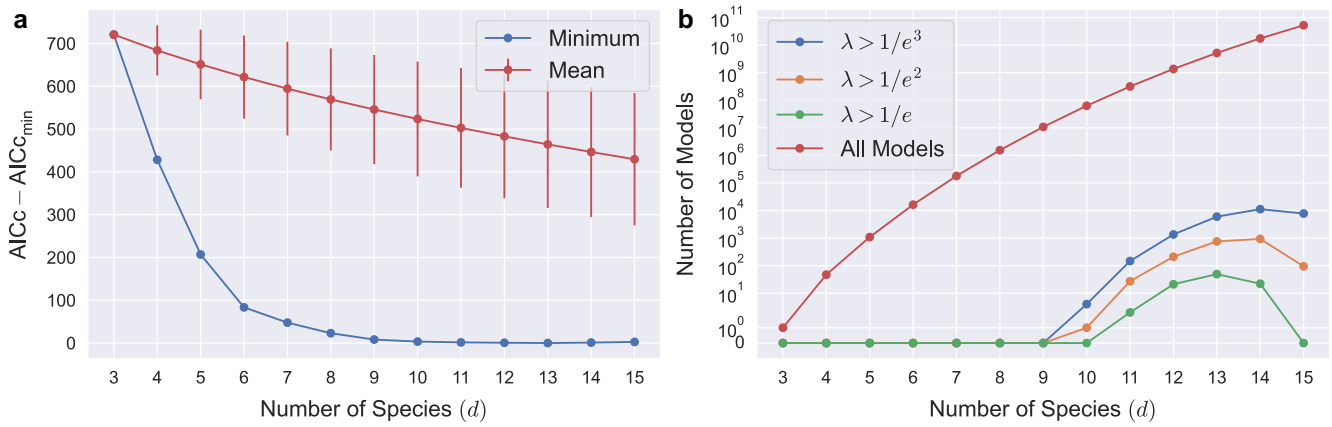
In the format provided by the authors and unedited

CONTENTS

I. SUPPLEMENTARY RESULTS

A. Modeling landscape

The exhaustive model search described in the main text consists of every possible species combination from $d = 3$ to 15 and covers nearly 80 billion possible plume models. However, most of these models do not accurately reproduce the observed INMS data. Owing to the combinatorially large parameter space, only $\sim$1 in 10^9 models exhibits strong predictive power. Supplementary Fig. 1 demonstrates that the average model performs drastically worse than the optimal model for each value of d. This performance difference is significant across multiple standard deviations. In other words, although each species in the spectral library is a plausible plume constituent a priori, a random selection of compounds from the library would very likely not be capable of fitting the INMS data. Interestingly, the mean AICc decreases monotonically over the entire parameter range while the minimum AICc increases from $d = 13$ to 15 (see also Fig. 2a of the main text). This trend demonstrates that the inclusion of additional species can appear to improve model performance indefinitely if only low likelihood models are considered. Such modeling behavior highlights the importance of exploring the entire parameter space in order to obtain accurate results (see also Supplementary Section II B).



Supplementary Fig. 1. Overview of the modeling landscape explored during the exhaustive model search. (a) The minimum (blue) and mean (red) AICc values across all models for each value of d species. Vertical bars indicate ± 1 standard deviation range. Note that the minimum AICc increases significantly from $d = 13$ to 15 but appears flat on this axis scale (see Fig. 2a of the main text). (b) The number of models with $\lambda > 1/e^3$ (blue), $\lambda > 1/e^2$ (orange), and $\lambda > 1/e$ (green) compared to the total number of models (red) for each value of d.

B. Alternative model selection criteria

As discussed in the main text, it is advisable to limit the total number of candidate models in order to make reliable statistical inferences. This consideration motivates the use of a restricted spectral library consisting of only geochemically plausible (and INMS detectable) compounds, as well as the elimination of extraneous low likelihood models. The criteria used to define which models are "high likelihood" requires a tradeoff between neglecting alternative explanations (i.e., bias; underfitting) and risking parameter estimation based on poor performing models (i.e., variance; overfitting). The inferences of the main text are based on the exhaustive list of models satisfying $\lambda > 1/e^3$. This threshold was chosen to reduce bias and because Monte Carlo simulations suggest that it corresponds to a $\sim$95–99% model confidence interval [S1–S4]. To assess the influence of this value on our results, we repeated the analysis of the main text for the increasingly restrictive thresholds $\lambda > 1/e^2$ and $\lambda > 1/e$. Correspondingly, results based on these criteria are made using an increasingly small set of only the highest likelihood models (see also Fig. 2a of the main text). In both cases, our results remain virtually unchanged (Supplementary Table 1). The use of a more restrictive threshold leads to slightly more precise mixing ratios for some minor compounds, as expected with a bias-variance tradeoff. Notably, even the lowest likelihood species listed in Table 1 of the main text are also contained within the more restricted sets of models. This result suggests that adopting the lower (more unbiased) threshold of $\lambda > 1/e^3$ does not lead to overfitting as a result of the increased variance.

Supplementary Table 1. Modeling results for alternative model selection criteria. Mean values and uncertainties are presented as in the main text.

	$\lambda > 1/e$		$\lambda > 1/e^2$		Reduced ($\lambda > 1/e^3$)	
Species	Probability	Mixing Ratio (%)	Probability	Mixing Ratio (%)	Probability	Mixing Ratio (%)
CH_4	1	0.11 ± 0.02	1	0.11 ± 0.02	1	0.11 ± 0.02
NH_3	1	1.8 ± 0.1	1	1.8 ± 0.1	1	1.8 ± 0.1
H_2O	1	95.9 ± 0.3	1	95.9 ± 0.3	1	95.9 ± 0.3
C_2H_2	1	0.025 ± 0.003	0.95	0.024 ± 0.004	0.78	0.020 ± 0.009
HCN	1	0.11 ± 0.02	1	0.11 ± 0.02	1	0.10 ± 0.02
C_2H_4	0	$< 6. \times 10^{-4}$[a]	0.06	< 0.03	0.22	< 0.1
CO	1	0.73 ± 0.06	1	0.73 ± 0.07	1	0.72 ± 0.07
N_2	0	$< 6 \times 10^{-4}$[a]	0.01	< 0.002	0	$< 6. \times 10^{-4}$[a]
CH_2O	0	< 0.001[a]	0.03	< 0.002	0.07	< 0.005
NO	0.01	$< 7. \times 10^{-4}$[a]	0.03	< 0.001	0.12	< 0.006
C_2H_6	0.81	0.016 ± 0.008	0.74	0.014 ± 0.009	0.57	0.011 ± 0.010
CH_5N	0.07	< 0.001	0.10	< 0.002	0.22	< 0.005
CH_3OH	0.49	< 0.005	0.48	< 0.005	0.42	< 0.005
H_2S	0.38	< 0.003	0.39	< 0.003	0.33	< 0.002
PH_3	0.19	< 0.001	0.22	< 0.002	0.18	< 0.001
O_2	0.59	< 0.007	0.60	< 0.008	0.49	< 0.007
^{36}Ar	0	$< 4. \times 10^{-4}$[a]	0.02	$< 4. \times 10^{-4}$[a]	0	$< 4. \times 10^{-4}$[a]
C_3H_4	0.05	$< 8. \times 10^{-4}$[a]	0.10	< 0.001	0.08	< 0.001
^{40}Ar	0.76	0.0012 ± 0.0010	0.66	0.0011 ± 0.0010	0.58	< 0.004
CH_3CN	0.07	< 0.001	0.11	< 0.002	0.07	< 0.002
C_2H_2O	0	$< 7. \times 10^{-4}$[a]	0.01	$< 7. \times 10^{-4}$[a]	0	$< 7. \times 10^{-4}$[a]
C_3H_6	1	0.005 ± 0.002	0.92	0.004 ± 0.002	0.94	0.005 ± 0.002
CH_2N_2	0	$< 4. \times 10^{-4}$[a]	0.04	$< 4. \times 10^{-4}$[a]	0.03	$< 4. \times 10^{-4}$[a]
C_2H_4O	0	$< 7. \times 10^{-4}$[a]	0.05	< 0.001	0.12	< 0.003
C_3H_8	0.19	< 0.005	0.22	< 0.005	0.22	< 0.006
CO_2	1	0.45 ± 0.03	1	0.45 ± 0.03	1	0.45 ± 0.03
C_2H_7N	0	< 0.002[a]	0.01	< 0.002[a]	0	< 0.002[a]
CH_3NO	0	$< 5. \times 10^{-4}$[a]	0.01	$< 5. \times 10^{-4}$[a]	0	$< 5. \times 10^{-4}$[a]
C_2H_6O	0	< 0.002[a]	0.01	< 0.002[a]	0	< 0.002[a]
CH_2O_2	0	$< 7. \times 10^{-4}$[a]	0.01	$< 7. \times 10^{-4}$[a]	0	$< 7. \times 10^{-4}$[a]
C_4H_8	0	$< 7. \times 10^{-4}$[a]	0.01	$< 7. \times 10^{-4}$[a]	0	$< 7. \times 10^{-4}$[a]
$C_2H_6N_2$	0.38	< 0.002	0.29	< 0.002	0.12	$< 9. \times 10^{-4}$
C_3H_6O	0.23	< 0.001	0.17	< 0.001	0.10	$< 7. \times 10^{-4}$
C_4H_{10}	0.04	< 0.008[a]	0.08	< 0.008[a]	0.07	< 0.008[a]
$C_2H_4O_2$	0	$< 6. \times 10^{-4}$[a]	0.01	$< 6. \times 10^{-4}$[a]	0	$< 6. \times 10^{-4}$[a]
C_3H_8O	0.06	< 0.003[a]	0.09	< 0.004	0.02	< 0.003[a]
C_2H_7NO	0.16	< 0.002	0.21	< 0.002	0.19	< 0.002
$C_2H_6O_2$	0.23	< 0.01	0.21	< 0.01	0.04	< 0.005[a]
C_5H_{10}	0	$< 6. \times 10^{-4}$[a]	0.01	$< 6. \times 10^{-4}$[a]	0	$< 6. \times 10^{-4}$[a]
C_4H_9N	0.02	$< 2. \times 10^{-4}$[a]	0.04	$< 2. \times 10^{-4}$[a]	0.05	$< 2. \times 10^{-4}$[a]
C_5H_{12}	0.02	< 0.002[a]	0.05	< 0.002[a]	0.05	< 0.002[a]
$C_4H_{10}O$	0	$< 4. \times 10^{-4}$[a]	0.02	$< 4. \times 10^{-4}$[a]	0.01	$< 4. \times 10^{-4}$[a]
$C_2H_5NO_2$	0.02	< 0.007	0.04	< 0.01	0.08	< 0.02
C_3H_5Cl	0.08	$< 4. \times 10^{-4}$	0.18	< 0.001	0.02	$< 2. \times 10^{-4}$[a]
C_5H_9N	0.02	$< 3. \times 10^{-4}$[a]	0.10	$< 3. \times 10^{-4}$[a]	0.03	$< 3. \times 10^{-4}$[a]
$C_4H_6O_2$	0.05	$< 6. \times 10^{-4}$[a]	0.07	$< 6. \times 10^{-4}$[a]	0.07	$< 6. \times 10^{-4}$[a]
$C_3H_7NO_2$	0	$< 1. \times 10^{-4}$[a]	0.02	$< 3. \times 10^{-4}$	0	$< 1. \times 10^{-4}$[a]
C_8H_{18}	0.02	$< 1. \times 10^{-4}$[a]	0.03	$< 1. \times 10^{-4}$[a]	0.04	$< 1. \times 10^{-4}$[a]
$C_6H_{12}N_4$	0	$< 5. \times 10^{-5}$[a]	0.03	$< 5. \times 10^{-5}$[a]	0	$< 5. \times 10^{-5}$[a]
$C_6H_{14}N_2O_2$	0	$< 1. \times 10^{-4}$[a]	0.01	$< 1. \times 10^{-4}$[a]	0	$< 1. \times 10^{-4}$[a]

[a] Denotes number density relative to H_2O.

The decision to use the exhaustive model list was made to further reduce model selection bias as much as possible. We believe this is appropriate given the limited data and large number of plausible model components [S5]. Nevertheless, alternative procedures could be implemented by factoring in prior knowledge about the INMS data set. To investigate the stability of our results under different model selection criteria, we repeated the analysis once more by applying the following additional selection criteria to the initial set of $\lambda > 1/e^3$ models. First, we excluded all models containing more than one species with mass greater than 50 u, above which INMS did not detect any signal above background noise [S6, S7]. Doing so conserves statistical power in an effort to focus on lower mass compounds for which there is prior circumstantial evidence [S8, S9]. Next, in order to reduce redundancy, we eliminated models with any pairwise regressor correlations greater than $\rho = 0.9$ [S1, S5, S10] (see also Supplementary Section I C below). Models with $d > 13$ were also excluded for the same purpose (see Fig. 2b of the main text). Finally, if there existed a simpler model with a lower AICc that was fully nested within a more complex model, the more complex model was discarded. In this scenario, the more complex model performs well only because it contains all the parameters of the simpler model, and some studies suggest that this rule improves model selection accuracy by reducing the chance of false positive results [S1–S3, S5]. The species probabilities and mixing ratios determined on the reduced set of models are shown in the rightmost column of Supplementary Table 1. Once again, there is very good agreement between these results and those of the main text, further indicating that the reduction in model bias imposed by the exhaustive search is not associated with overfitting.

Overall, the strong similarity of the results across multiple different selection procedures demonstrates stability in the modeling process. Nevertheless, we recommend that quantitative inferences be made using the values presented in the main text. Those results represent the most unbiased (and conservative) determination of the plume's composition. By contrast, the increased precision of the results listed in Supplementary Table 1 may not be warranted and should be interpreted only as supporting evidence.

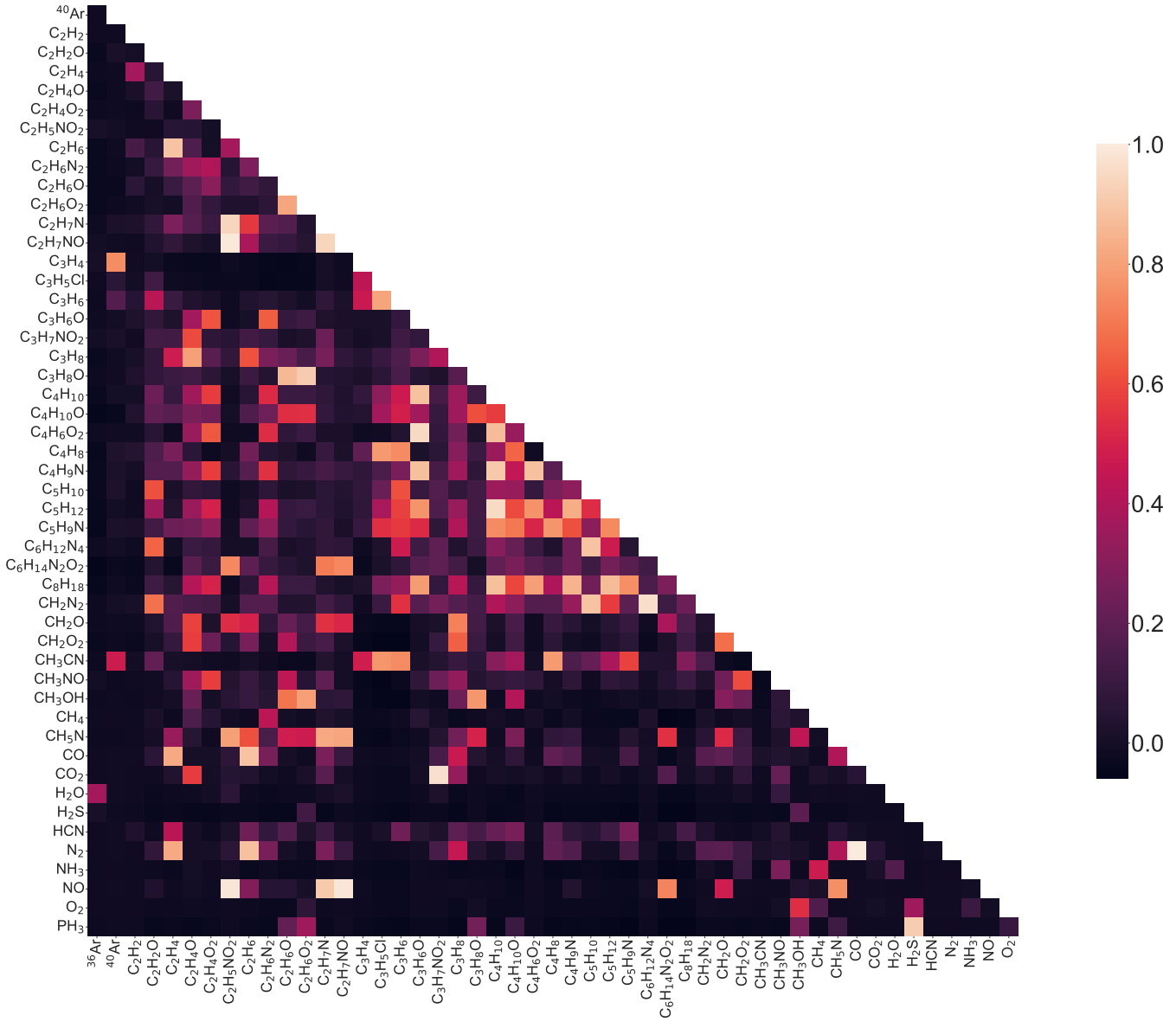
C. Analysis of pairwise collinearity

Compositional ambiguities within INMS spectra arise due to the large number of candidate species combined with the relatively low mass resolution of the instrument. In other words, models of the plume's composition contain many possible model components with comparatively few data points available to constrain them. This difficulty is accentuated when there exist combinations of multiple species that can reproduce the signal produced by another. If this collinearity is exact, discriminating between such singular species in the composite INMS spectrum is impossible. In practice, even approximately collinear mass spectra can present a significant obstacle when interpreting data with a finite mass resolution. These issues have been encountered in previous studies [S7–S9] and have greatly hindered efforts to identify trace compounds in the plume.

Here, we quantify the extent of pairwise collinearity amongst library spectra by computing the correlation matrix for each species' cracking pattern. We use ρ to denote correlation coefficients of cracking patterns(i.e., the regressors themselves), in contrast to r (defined in the main text), which signifies correlations between regression coefficients. Whereas r is a model dependent quantity (see Fig. 2b of the main text), ρ is a static property of the spectral library.

Supplementary Fig. 2 demonstrates that large positive correlations are common and manifest between molecules with similar cracking patterns. A correlation coefficient between two species of $\rho = 1$ would imply identical mass spectra. Pairs of species with correlation coefficients close to 1 are approximately collinear and may still be indistinguishable in practice. By contrast, pairs of species with correlation coefficients close to 0 lack a consistent pattern of strong overlapping features in their cracking patterns. Values near 0.5 represent the intermediate case where two species share certain features but also possess additional large mass peaks that are not shared. NH_3 and CH_4, for example, have a correlation coefficient of 0.47, owing to their shared peak at mass 16 and unshared peaks at masses 17 (NH_3) and 15 (CH_4). Of course, large negative values are also possible, though they are effectively absent from the spectral library due to the inherent structural similarities between organic compounds. Large negative values would signify that one species has significant mass peaks predominantly at mass channels where another species does not.

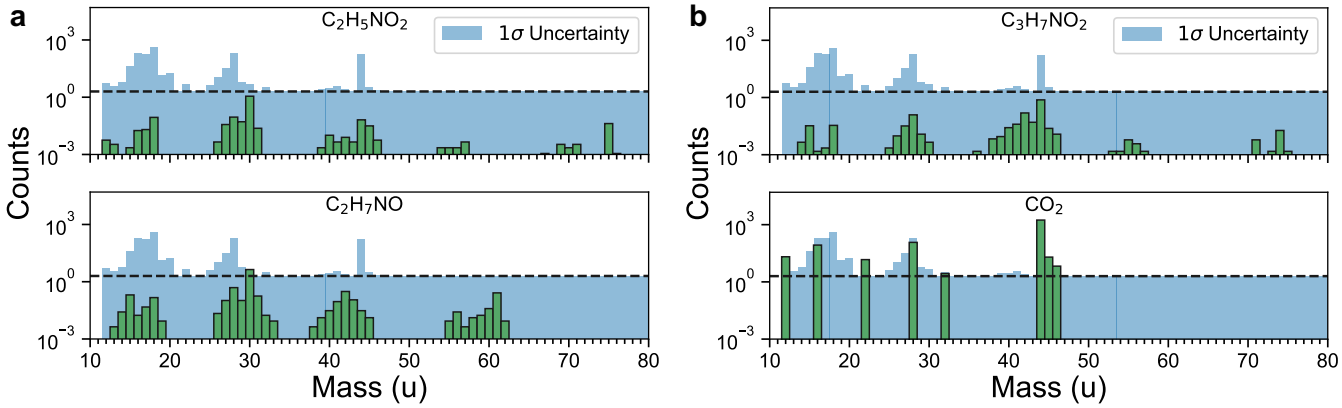
For compounds with cracking patterns dominated by only a few major peaks, sharing as few as one of these peaks can lead to high correlation coefficients. A notable example is C_2H_7NO and glycine ($C_2H_5NO_2$), which exhibit a correlation coefficient of $\rho = 0.99$. Based on the results of the main text, it is tempting to view the presence of glycine in 2 of the 94 highest likelihood models ($\lambda > 1/e$) as the first tentative evidence for amino acids at Enceladus. However, at such low concentrations, glycine is virtually indistinguishable from C_2H_7NO. Although the cracking pattern for glycine contains counts at many different mass channels, the peak at 30 u is by far the dominant feature. As a result, this mass channel acts as a high leverage point and drags up correlations between glycine and other species with large peaks at 30 u. This effect is particularly pronounced when there are no other major peaks present, as is the case for C_2H_7NO. Supplementary Fig. 3a shows the mean contributions of glycine and C_2H_7NO to the INMS spectrum calculated based on the multi-model averaging procedure presented in the main text. All peaks are well within the



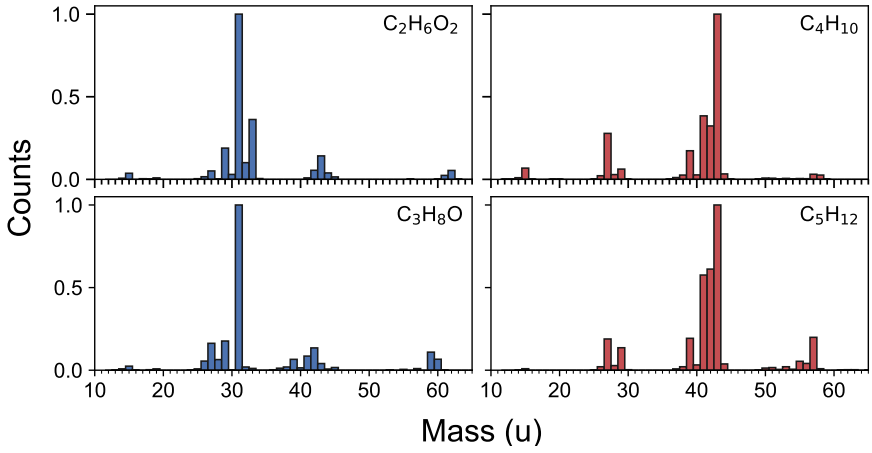
Supplementary Fig. 2. Correlation matrix for the spectral library. The correlation coefficient ($-1 \leq \rho \leq 1$) measures the extent of collinearity between two cracking patterns. Large positive values (brighter colors) indicate high levels of collinearity, whereas smaller values signify less collinearity. Large negative values (darker colors) are essentially absent, indicating that most collinear species are positively correlated.

associated 1σ uncertainty for each mass channel. The base peak is the only C_2H_7NO feature that extends above the minimum count uncertainty and mimics the signal for glycine at similar concentrations. This ambiguity precludes the detection of trace amounts of glycine in the plume and might have important implications for the detection of amino acids on future spacecraft missions as well. The high pairwise correlation between glycine and C_2H_7NO suggests that glycine likely cannot be independently detected at Enceladus using a 1 u resolution mass spectrometer (such as INMS). Instead, an independent measurement of the C_2H_7NO mixing ratio with precision at least as great as the instrument used to detect glycine would be necessary. A similar argument applies to alanine ($C_3H_7NO_2$) and CO_2 ($\rho = 0.97$) which share a single dominant peak at mass 44 (Supplementary Fig. 3b).

Ambiguities of this nature also underlie the increased mixing ratio uncertainties deduced for the alcohols and mass 43 fragments discussed in the main text. Large positive correlations amongst the alcohols (as high as $\rho = 0.91$) manifest via high count rates at 31 u, corresponding to the hydroxymethyl group, CH_2OH^+. Species with prominent 43 u fragments exhibit correlations as high as $\rho = 0.95$ and could be attributable to a number of different structures (see Supplementary Fig. 4 and Supplementary Section II A). Consequently, inter-model uncertainty stemming from



Supplementary Fig. 3. (a) Mass spectra for glycine and C_2H_7NO. The x-axis shows the mass of each fragmentation product, while the y-axis quantifies the proportional abundance of each fragment based on the mean mixing ratios determined in the main text (note the log-scale). All spectral features (green bars) are masked by the associated count uncertainty at each mass channel (shaded blue region). Both species have a single dominant peak at 30 u, leading to high pairwise correlations. Only the signal from C_2H_7NO extends above the minimum count uncertainty (dashed black line). (b) Mass spectra for alanine and CO_2. Both species share a single dominant peak at mass 44.



Supplementary Fig. 4. Mass spectra for representative alcohols and species with 43 u fragments. Blue bars (left) show representative alcohols, $C_2H_6O_2$ and C_3H_8O, with $\rho = 0.91$. Red bars (right) show representative species with strong 43 u signatures, C_4H_{10} and C_5H_{12}, with $\rho = 0.95$.

these correlations reduces the precision of the final averaged mixing ratio.

D. Ramifications for hypothesis testing

Extensive amounts of collinearity between model components can have profound consequences on statistical inference. Although one might expect traditional hypothesis testing to identify important model parameters, high dimensionality and collinearity of the spectral library limit statistical power and prevent individual regression coefficients from achieving statistical significance. Supplementary Table 2 demonstrates this phenomenon for the low velocity INMS data set. An F-test for overall significance indicates that the spectral library, in aggregate, does indeed better explain the INMS data than does the "null model," which consists of fitting only the regression coefficient for H_2O ($F = 12$; $p = 1.8 \times 10^{-14}$). However, one-tailed t-tests indicate that only the three strongest spectral features, H_2O, NH_3, and CO_2 are individually statistically significant ($p = 2.6 \times 10^{-79}$, 6.0×10^{-16}, and 9.5×10^{-7}, respectively) in the presence of the entire spectral library. These species account for the majority of counts at masses 16, 17, 18, and 44 and produce signals that are large enough to stand out amongst the majority of collinear species that comprise the rest of the library. Though we can be confident that the aggregate set of candidate library species is capable of

explaining the INMS data, the statistical significance of any one species is difficult to show using such frequentist statistics under these conditions. For this reason, information theoretic approaches (such as those described in the main text) or related Bayesian analyses are more suitable for INMS spectra deconvolution.

Supplementary Table 2. Results of hypothesis testing on the INMS data set. High dimensionality and collinearity prevent all but the most prominent spectral features from achieving statistical significance.

Species	t-Statistic	p-Value
H_2O	380	2.6×10^{-79}
NH_3	12	6.0×10^{-16}
CO_2	5.5	9.5×10^{-7}
All others	< 1.4	> 0.08

II. SUPPLEMENTARY DISCUSSION

A. Comparison to other results

It is important to consider the entire body of statistical evidence when drawing conclusions about which species are detected in the plume. The model-averaged mixing ratios, the minimum AICc model, the individual species probabilities, and the relative model likelihoods can all be used to assess the confidence of each detection. The more heuristics that point towards the detection of a given species, the more confident one may be that said species is truly present in the plume.

The minimum AICc model consists of 13 species. One interpretation of this result is to classify all 13 species as conclusive detections. However, more parsimonious models exist that explain the INMS data nearly as well. A more conservative approach would be to interpret only the species that comprise the best-fitting, least complex model with $\lambda > 1/e^3$ as essential to the fit. We favor an even more nuanced approach and suggest using a tiered hierarchy of confidence based on the holistic analysis of the main text. Below, we contextualize these results within the collection of previous studies on the composition of Enceladus and other icy bodies.

In the main text, we present an NH_3 mixing ratio that is larger than what has been reported based on previous analyses of the slow flyby INMS spectra [S6, S7], as well as those obtained during the E2 [S8] and E3 [S9] flybys. In their supplementary material, Waite et al. [S6] argue, based on the slow flyby data, that the residual left at mass 16 by fitting H_2O, CH_4, and CO_2 alone unambiguously indicates the presence of NH_3. For the spectrum analyzed in the main text of this work, we find that this residual is equal to ∼1330 counts. The NH_3 / H_2O number density calculated based on this residual alone is ∼0.01, in agreement with ref. [S6]. Of course, this conclusion is dependent on the estimation of count uncertainties in INMS data, for which multiple different procedures have been proposed [S6, S9, S11, S12]. The precise uncertainties adopted in ref. [S6] may not be identical to those used here, and this difference could be a source of diverging results. Telescopic data have suggested the presence of NH_3 or NH_3-hydrate [S13–S16] at Enceladus, though these observations are not definitive [S17] and other instruments including the *Cassini* Visible and Infrared Mapping Spectrometer (VIMS), Ultraviolet Imaging Spectrograph (UVIS), and Cosmic Dust Analyzer (CDA) have failed to find conclusive evidence for its presence in the plume [S7, S18, S19]. Although we agree with the authors of ref. [S6] that NH_3 is required to explain the INMS data, additional studies that further examine the effects of uncertainty estimation would be helpful in validating the precise mixing ratio.

As discussed in the main text, suggestive evidence for nitrogen at Enceladus in the form of HCN has been previously reported based on other INMS data sets. HCN has also been suggested to help explain yet unresolved signatures in Cassini CDA spectra of ice grains in Saturn's E-ring [S20], though could not be definitively identified by CDA as a cation species due to its high ionization potential [S21]. As such, the HCN mixing ratio determined in this work is the first definitive detection of nitrile chemistry at Enceladus.

Our analysis shows that C_2H_2 and C_2H_6 are also present in the plume with concentrations exceeding 100 ppm. As for HCN, both species have been detected in comets [S22] and circumstantial evidence for their existence at Enceladus has been previously suggested based on other Cassini flybys. C_2H_2 was suspected by Waite et al. [S9] during the high-velocity E3 and E5 flybys, though correlations between abundance and spacecraft velocity suggest that impact fragmentation may have been responsible. Only upper limits for C_2H_6 were reported based on the E2, E3, and E5 flybys [S9]. The presence of C_2H_6 would also help explain features at 28–30 u seen in CDA ice grain spectra [S7], which are consistent with the INMS data.

Concerning the higher mass organics, we find that trace amounts (∼40 ppm) of C_3H_6 (and possibly C_3H_8) are present in the plume as well. C_3H_6 was produced as a fragmentation product during the E3 and E5 flybys of Enceladus [S9], and both species have been observed at Titan [S11]. Although not initially identified in the E2 flyby data [S8], reanalysis suggests that C_3H_6 may have been present [S9]; however it was not identified by Waite et al. as an intrinsic plume constituent in subsequent publications [S6, S7]. Our analysis shows that C_3H_6 is by far the most likely explanation for the count signal in the 37–42 u region of the slow flyby INMS spectrum (Fig. 4d of the main text) and provides the first conclusive identification of native C_3 organics in the plume.

Native alcohols have not been previously identified in the plume, but CH_3OH has been suggested as a possible fragmentation product based on high velocity INMS and ice grain spectra [S23]. CH_3OH has also been observed via ground-based methods in the vicinity of Enceladus [S24] and has been proposed as an explanation for the 3.53 μm absorption feature seen in VIMS spectra of the surface [S25]. CH_3OH could be produced through chemical processing of CH_4 in the nearby gas cloud or by the partial combustion of endogenous CH_4 within the ocean. Both CH_3OH and $C_2H_6O_2$ have been detected in relatively high abundance in several comets [S26, S27].

As discussed in the main text, the authors of ref. [S6] identified a possible instrument artifact at mass 32 that complicates the determination of the native plume O_2 abundance. In accounting for this artifact, they estimated a corrected mass 32 signal equal to $(100/45) \times 0.004\% = 0.0089\%$ of the counts measured at mass 18. The value of mass 32 counts used in this work is well within this limit, and the concomitant detections of multiple partially oxidized organics are consistent with a native source of O_2. However, the precise source of O_2 at Enceladus presents a few challenges. Unlike Europa, where charged particle bombardment of the surface is known to drive radiolysis of water and other elements to O_2, H_2O_2, SO_4^{2-}, and other oxidants [S28–S31], the radiation flux near Enceladus is considerably lower [S32], and evidence for oxidants on the surface is lacking, despite proposals of such radiolytic chemistry [S33–S35]. Even with moderate levels of surface radiolysis, a key problem would be the efficiency of oxidant production at the high temperatures observed in the South Polar Terrain [S36, S37]. Here we do not propose a solution for this issue but rather note that our results are consistent with the presence of O_2, be it from surface radiolysis and subsequent delivery to the ocean, or via other production mechanisms. Notably, Waite et al. [S6] found that radiolysis of H_2O due to radioactive isotopes in Enceladus' core could also produce O_2 at the abundance reported here.

Evidence for ^{40}Ar stems from the large $\sim 2.2\sigma$ residual at 40 u, which is the only mass channel with strong enough signal to significantly influence the calculation of its mixing ratio. This signal does have strong overlap with the cracking pattern of C_3H_4 ($\rho = 0.75$), but the abundance of this species is limited by the larger contribution of C_3H_6 at neighboring mass channels. Though not identified in a previous analysis of the slow flybys [S6], ^{40}Ar was detected during the E3 and E5 flybys and may indicate significant water-rock interactions and the leaching of salts within Enceladus [S9, S38]. CH_3CN also has a reasonably high correlation with ^{40}Ar ($\rho = 0.47$) and could potentially contribute to the observed signal at mass 40. Although there is no definitive evidence for CH_3CN at Enceladus, its presence would not be unexpected based on the evidence for HCN and hydrocarbons such as C_3H_6 (see Discussion in the main text).

Our results also provide strong evidence for an ambiguous species with a strong 43 u signal. Ambiguity at this mass channel was previously reported based on the E5 flyby of Enceladus [S9]. One explanation of this signal is $C_2H_6N_2$. Although this species has an additional large peak at 15 u, this feature is masked by the much larger contribution from CH_4 in the INMS data. The correlated ice grain features at 15 and 43 u seen in "Type II" CDA spectra [S21] might be explained by fragmentation of $C_2H_6N_2$ into CH_3^+ and $CH_3N_2^+$. Alternative explanations including acetyl group-bearing species such as C_3H_6O or $C_4H_6O_2$ are also possible.

Sulfur compounds have not been definitively identified at Enceladus, though a tentative past detection of H_2S based on the E5 flyby [S7, S9] supports our finding that it may be present. H_2S would be expected if there is active serpentinization taking place on the ocean floor. PH_3 has not been previously reported in the plume, but may have been observed in the coma of comet 67P/Churyumov-Gerasimenko [S39]. A recent analysis of salt-rich ice grains collected by CDA in Saturn's E-ring detected phosphorus in the form of Na_3PO_4 and Na_2HPO_4, which may imply an abundant source of phosphorus within Enceladus' ocean [S40].

For the remaining species listed in Table 1 and Extended Data Table 2 of the main text, prior evidence for their existence in the plume is relatively poor. Neither CH_5N nor C_3H_5Cl have been identified at Enceladus, but Cl has been detected by CDA as NaCl and KCl salts residing in plume ice grains [S41, S42]. Lastly, although there is no strong evidence for amino acids at Enceladus (see Supplementary Section I C), we note that glycine, alanine, and other amino acids are abundant in carbonaceous chondrites [S43].

B. Comparison to other methodologies

In order to adequately account for the high dimensionality and (approximate) collinearity of the INMS plume dataset, it is necessary to perform a type of variable selection that constrains the parameter space of possible model

fits. Such variable selection techniques trade off a small increase in model bias for a significant reduction in model variability [S10, S44]. The resulting models are far less likely to overfit noisy features in the training data and tend to be significantly more accurate in predicting future observations [S10]. Moreover, variable selection reduces the impact of collinearity by identifying which model components are better at explaining the observed data and discarding those that are superfluous. Such a process allows for the evaluation of individual model components without the confounding presence of their collinear counterparts. Subsequent multi-model inference based on model averaging then ensures that all candidate models are considered on equal footing.

In the main text, we outlined two variable selection procedures: an exhaustive all subset selection for models with fewer than 16 species and a forward stepwise selection algorithm for more complex models. Additional model selection criteria are considered in Supplementary Section I B above. Other common algorithms such as ridge regression [S45] and the Least Absolute Shrinkage and Selection Operator (LASSO) are frequently applied in a wide variety of machine learning and model validation contexts [S10, S44, S46–S48]. These methods seek to reduce the influence of extraneous parameters via L2 or L1 regularization, respectively. Multi-model averaging is similar to L1 regularization in that it allows for explicit dimensionality reduction, whereas L2 regularization does not. This property leads to high model interpretability, which is of utmost importance when performing compositional analyses. Other heuristics besides the AICc can be used to select models for averaging, but these alternative statistics are not based on minimizing information loss and are therefore not well-suited for model selection when the structure of the unknown, true distribution is poorly constrained. The property of the AICc as a minimum variance unbiased estimator of model performance supports its use in this context [S49]. Furthermore, model inference using the AICc has been shown to asymptotically approximate results based on cross-validation (another broadly accepted model validation technique) while requiring much fewer computational resources [S1, S50, S51].

The first few studies of INMS data collected at Enceladus produced landmark results, including the characterization of major plume constituents and the detection of molecular H_2 as a potential indicator of hydrothermal activity [S6, S8, S9]. These papers (and related works published throughout the duration of the *Cassini* mission) also documented a detailed description of the INMS instrument response under varying spacecraft conditions and laid the groundwork for follow-up studies focused solely on compositional analyses. However, early studies of the Enceladus plume—though foundational—may not have been well-suited to resolve minor species ambiguities for various reasons. In order to facilitate comparison with our methodology, we briefly describe the spectral deconvolution procedure developed by the authors of ref. [S11] that has been implemented in various other works (e.g., refs. [S52–S54]).

In their procedure, the authors first determine the contributions from major species through a visual analysis of prominent spectral features. Mixing ratios for the major species are estimated from the base peaks of each species, assuming they contribute 100% of the measured counts at these mass channels. Minor species are then identified sequentially by subtracting their contributions from the total spectrum to produce a residual spectrum. For small portions of the spectrum where a few candidate species exhibit overlaying signatures, species are fit based on a custom-defined fit statistic using a grid search algorithm. Iterative minimization of the fit statistic is achieved numerically by sweeping through various mixing ratios at increasingly finer resolution. For species that share the same base peak (e.g., N_2 and C_2H_4), the fit statistic is manipulated to exclude this mass channel. The high computational intensity of the grid search algorithm prohibits fitting more than four species at a time.

We believe that the methodology of ref. [S11] described above, though useful and effective, may not be optimal for identifying minor species in the INMS data. The order in which species are subtracted from the initial spectrum will generally influence the outcome of the analysis (see the discrepancy between the exhaustive and forward modeling approaches presented in the main text). Although it is true that a bias-variance tradeoff can be useful for combatting high dimensionality, the procedure of ref. [S11] is not amenable to quantitative assessments of inter-model uncertainty. Indeed, the authors note that subjectivity of their analysis is a valid concern. Furthermore, the practice of fitting individual species to small portions of the spectrum neglects potential contributions from complex compounds with cracking patterns that span a large mass range. Moreover, grid searches that consider only a few species at a time may not be able to reliably identify minor species when the spectral library is highly collinear. In this regime, covariances between mixing ratios become strongly model dependent (see Fig. 2 of the main text), and multi-model inference based on model-averaged parameters is warranted [S3]. Additionally, the fit statistic described in ref. [S11] is not a standard metric, and we suspect that the use of different fit statistics for different model parameters could lead to difficulties in interpreting the results. Lastly, the authors' procedure does not employ dimensionality reduction or account for the possibility of over-fitting to noise.

By contrast, our approach quantitatively addresses both inter- and intra-model uncertainty in the spectral decomposition of INMS data. While previous studies, such as those presented in refs. [S7, S55], have concluded that minor species identification requires a higher resolution mass spectrometer, we have presented a mathematical framework capable of discriminating between previously ambiguous species. The heuristics used in this analysis are based on maximum likelihood estimation and relative entropy minimization—foundational principles of statistical inference and information theory.

Nevertheless, this study is not without limitations. A major challenge for any compositional analysis of INMS data stems from the large number of candidate plume species. The chemistry of the ocean and ice shell could include hundreds to thousands of unique compounds that contribute to the observed INMS spectrum. Our approach using the AICc is based on the principle of parsimony in that the least-complex, best-fitting model is favored over similarly performing models of higher complexity. Although this is a fruitful approach to developing conservative models of the plume's composition, nature does not necessarily reflect this ideal. Future investigations using higher resolution mass spectrometers with larger mass ranges will shed light on the full extent of chemical diversity within the plume and the ocean beneath.

A number of interesting follow-up studies could be conducted to validate the results presented in this work. A strong approach would be to treat each of the slow flybys (E14, E17, and E18) as individual data sets, as opposed to averaging them together. Machine learning models could then be trained on one data set and evaluated on another. A sort of round-robin procedure could be used to estimate the uncertainty associated with training on a particular data set. Such a methodology would eliminate the need for heuristic statistics such as the AICc in favor of actual independent test set performance. This implementation would, however, require correcting for instrument artifacts in each of the individual Enceladus flybys.

[S1] K. P. Burnham and D. R. Anderson, eds., *Model Selection and Multimodel Inference* (Springer New York, New York, NY, 2004).

[S2] S. A. Richards, Dealing with overdispersed count data in applied ecology: Overdispersed count data, Journal of Applied Ecology **45**, 218 (2007).

[S3] S. A. Richards, M. J. Whittingham, and P. A. Stephens, Model selection and model averaging in behavioural ecology: the utility of the IT-AIC framework, Behavioral Ecology and Sociobiology **65**, 77 (2011).

[S4] S. A. Richards, TESTING ECOLOGICAL THEORY USING THE INFORMATION-THEORETIC APPROACH: EXAMPLES AND CAUTIONARY RESULTS, Ecology **86**, 2805 (2005).

[S5] M. R. E. Symonds and A. Moussalli, A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion, Behavioral Ecology and Sociobiology **65**, 13 (2011).

[S6] J. H. Waite, C. R. Glein, R. S. Perryman, B. D. Teolis, B. A. Magee, G. Miller, J. Grimes, M. E. Perry, K. E. Miller, A. Bouquet, J. I. Lunine, T. Brockwell, and S. J. Bolton, Cassini finds molecular hydrogen in the Enceladus plume: Evidence for hydrothermal processes, Science **356**, 155 (2017).

[S7] F. Postberg, R. N. Clark, C. J. Hansen, A. J. Coates, C. M. Dalle Ore, F. Scipioni, M. M. Hedman, and J. H. Waite, Plume and Surface Composition of Enceladus, in *Enceladus and the Icy Moons of Saturn*, edited by P. Schenk (University of Arizona Press, Tucson, 2018) pp. 129–162.

[S8] J. H. Waite, M. R. Combi, W.-H. Ip, T. E. Cravens, R. L. McNutt, W. Kasprzak, R. Yelle, J. Luhmann, H. Niemann, D. Gell, B. Magee, G. Fletcher, J. Lunine, and W.-L. Tseng, Cassini Ion and Neutral Mass Spectrometer: Enceladus Plume Composition and Structure, Science **311**, 1419 (2006).

[S9] J. H. Waite Jr, W. S. Lewis, B. A. Magee, J. I. Lunine, W. B. McKinnon, C. R. Glein, O. Mousis, D. T. Young, T. Brockwell, J. Westlake, M.-J. Nguyen, B. D. Teolis, H. B. Niemann, R. L. McNutt Jr, M. Perry, and W.-H. Ip, Liquid water on Enceladus from observations of ammonia and 40Ar in the plume, Nature **460**, 487 (2009).

[S10] G. James, D. Witten, T. Hastie, and R. Tibshirani, eds., *An introduction to statistical learning: with applications in R*, Springer texts in statistics No. 103 (Springer, New York, 2013).

[S11] B. A. Magee, J. H. Waite, K. E. Mandt, J. Westlake, J. Bell, and D. A. Gell, INMS-derived composition of Titan's upper atmosphere: Analysis methods and model comparison, Planetary and Space Science **57**, 1895 (2009).

[S12] J. Cui, R. V. Yelle, V. Vuitton, J. H. Waite, W. T. Kasprzak, D. A. Gell, H. B. Niemann, I. C. F. Müller-Wodarg, N. Borggren, G. G. Fletcher, E. L. Patrick, E. Raaen, and B. A. Magee, Analysis of Titan's neutral upper atmosphere from Cassini Ion Neutral Mass Spectrometer measurements, Icarus **200**, 581 (2009).

[S13] M. Zastrow, J. T. Clarke, A. R. Hendrix, and K. S. Noll, UV spectrum of Enceladus, Icarus **220**, 29 (2012).

[S14] W. Grundy, Near-Infrared Spectra of Icy Outer Solar System Surfaces: Remote Determination of H2O Ice Temperatures, Icarus **142**, 536 (1999).

[S15] J. P. Emery, D. M. Burr, D. P. Cruikshank, R. H. Brown, and J. B. Dalton, Near-infrared (0.8–4.0 $\mathsf{\mu}$m) spectroscopy of Mimas, Enceladus, Tethys, and Rhea, Astronomy & Astrophysics **435**, 353 (2005).

[S16] A. J. Verbiscer, D. E. Peterson, M. F. Skrutskie, M. Cushing, P. Helfenstein, M. J. Nelson, J. Smith, and J. C. Wilson, Near-infrared spectra of the leading and trailing hemispheres of Enceladus, Icarus **182**, 211 (2006).

[S17] D. Cruikshank, T. Owen, C. Ore, T. Geballe, T. Roush, C. Debergh, S. Sandford, F. Poulet, G. Benedix, and J. Emery, A spectroscopic study of the surfaces of Saturn's large satellites: HO ice, tholins, and minor constituents, Icarus **175**, 268 (2005).

[S18] R. H. Brown, R. N. Clark, B. J. Buratti, D. P. Cruikshank, J. W. Barnes, R. M. E. Mastrapa, J. Bauer, S. Newman, T. Momary, K. H. Baines, G. Bellucci, F. Capaccioni, P. Cerroni, M. Combes, A. Coradini, P. Drossart, V. Formisano, R. Jaumann, Y. Langevin, D. L. Matson, T. B. McCord, R. M. Nelson, P. D. Nicholson, B. Sicardy, and C. Sotin, Composition and Physical Properties of Enceladus' Surface, Science **311**, 1425 (2006).

[S19] C. Hansen, L. Esposito, J. Colwell, A. Hendrix, G. Portyankina, A. Stewart, and R. West, The composition and structure of Enceladus' plume from the complete set of Cassini UVIS occultation observations, Icarus **344**, 113461 (2020).

[S20] J. K. Hillier, S. F. Green, N. McBride, J. P. Schwanethal, F. Postberg, R. Srama, S. Kempf, G. Moragas-Klostermeyer, J. A. M. McDonnell, and E. Grun, The composition of Saturn's E ring, Monthly Notices of the Royal Astronomical Society **377**, 1588 (2007).

[S21] F. Postberg, S. Kempf, J. K. Hillier, R. Srama, S. F. Green, N. McBride, and E. Grün, The E-ring in the vicinity of Enceladus: II. Probing the moon's interior—The composition of E-ring particles, Icarus **193**, 438 (2008).

[S22] M. J. Mumma and S. B. Charnley, The Chemical Composition of Comets—Emerging Taxonomies and Natal Heritage, Annual Review of Astronomy and Astrophysics **49**, 471 (2011).

[S23] F. Postberg, N. Khawaja, B. Abel, G. Choblet, C. R. Glein, M. S. Gudipati, B. L. Henderson, H.-W. Hsu, S. Kempf, F. Klenner, G. Moragas-Klostermeyer, B. Magee, L. Nölle, M. Perry, R. Reviol, J. Schmidt, R. Srama, F. Stolz, G. Tobie, M. Trieloff, and J. H. Waite, Macromolecular organic compounds from the depths of Enceladus, Nature **558**, 564 (2018).

[S24] E. Drabek-Maunder, J. Greaves, H. J. Fraser, D. L. Clements, and L.-N. Alconcel, Ground-based detection of a cloud of methanol from Enceladus: when is a biomarker not a biomarker?, International Journal of Astrobiology **18**, 25 (2019), publisher: Cambridge University Press.

[S25] R. Hodyss, C. D. Parkinson, P. V. Johnson, J. V. Stern, J. D. Goguen, Y. L. Yung, and I. Kanik, Methanol on Enceladus, Geophysical Research Letters **36**, L17103 (2009).

[S26] N. Biver, D. Bockelée-Morvan, V. Debout, J. Crovisier, J. Boissier, D. C. Lis, N. Dello Russo, R. Moreno, P. Colom, G. Paubert, R. Vervack, and H. A. Weaver, Complex organic molecules in comets C/2012 F6 (Lemmon) and C/2013 R1 (Lovejoy): detection of ethylene glycol and formamide, Astronomy & Astrophysics **566**, L5 (2014).

[S27] J. Crovisier, D. Bockelée-Morvan, N. Biver, P. Colom, D. Despois, and D. C. Lis, Ethylene glycol in comet C/1995 O1 (Hale-Bopp), Astronomy & Astrophysics **418**, L35 (2004).

[S28] K. P. Hand, *On the physics and chemistry of the ice shell and sub-surface ocean of Europa*, Ph.D. thesis, Stanford University (2007).

[S29] K. P. Hand and M. E. Brown, KECK II OBSERVATIONS OF HEMISPHERICAL DIFFERENCES IN H $_2$ O $_2$ ON EUROPA, The Astrophysical Journal **766**, L21 (2013).

[S30] J. R. Spencer and W. M. Calvin, Condensed O[TINF]2[/TINF] on Europa and Callisto, The Astronomical Journal **124**, 3400 (2002).

[S31] S. K. Trumbo, M. E. Brown, and K. P. Hand, H $_2$ O $_2$ within Chaos Terrain on Europa's Leading Hemisphere, The Astronomical Journal **158**, 127 (2019).

[S32] C. Paranicas, E. Roussos, N. Krupp, P. Kollmann, A. Hendrix, T. Cassidy, R. Johnson, P. Schenk, G. Jones, J. Carbary, D. Mitchell, and K. Dialynas, Energetic charged particle weathering of Saturn's inner satellites, Planetary and Space Science **61**, 60 (2012).

[S33] C. Ray, C. R. Glein, J. H. Waite, B. Teolis, T. Hoehler, J. A. Huber, J. Lunine, and F. Postberg, Oxidation processes diversify the metabolic menu on Enceladus, Icarus **364**, 114248 (2021).

[S34] S. F. Newman, B. J. Buratti, R. Jaumann, J. M. Bauer, and T. W. Momary, Hydrogen Peroxide on Enceladus, The Astrophysical Journal **670**, L143 (2007).

[S35] M. J. Loeffler and R. A. Baragiola, IS THE 3.5 um INFRARED FEATURE ON ENCELADUS DUE TO HYDROGEN PEROXIDE?, The Astrophysical Journal **694**, L92 (2009).

[S36] K. P. Hand and R. W. Carlson, H2O2 production by high-energy electrons on icy satellites as a function of surface temperature and electron flux, Icarus **215**, 226 (2011).

[S37] J. R. Spencer, J. C. Pearl, M. Segura, F. M. Flasar, A. Mamoutkine, P. Romani, B. J. Buratti, A. R. Hendrix, L. J. Spilker, and R. M. C. Lopes, Cassini Encounters Enceladus: Background and the Discovery of a South Polar Hot Spot, Science **311**, 1401 (2006).

[S38] C. R. Glein, F. Postberg, and S. D. Vance, The Geochemistry of Enceladus: Composition and Controls, in *Enceladus and the Icy Moons of Saturn* (The University of Arizona Press, 2018).

[S39] K. Altwegg, H. Balsiger, A. Bar-Nun, J.-J. Berthelier, A. Bieler, P. Bochsler, C. Briois, U. Calmonte, M. R. Combi, H. Cottin, J. De Keyser, F. Dhooghe, B. Fiethe, S. A. Fuselier, S. Gasc, T. I. Gombosi, K. C. Hansen, M. Haessig, A. Jäckel, E. Kopp, A. Korth, L. Le Roy, U. Mall, B. Marty, O. Mousis, T. Owen, H. Rème, M. Rubin, T. Sémon, C.-Y. Tzou, J. Hunter Waite, and P. Wurz, Prebiotic chemicals—amino acid and phosphorus—in the coma of comet 67P/Churyumov-Gerasimenko, Science Advances **2**, e1600285 (2016).

[S40] F. Postberg, Y. Sekine, F. Klenner, C. R. Glein, Z. Zou, B. Abel, K. Furuya, J. K. Hillier, N. Khawaja, S. Kempf, L. Noelle, T. Saito, J. Schmidt, T. Shibuya, R. Srama, and S. Tan, Detection of phosphates originating from Enceladus's ocean, Nature **618**, 489 (2023).

[S41] F. Postberg, S. Kempf, J. Schmidt, N. Brilliantov, A. Beinsen, B. Abel, U. Buck, and R. Srama, Sodium salts in E-ring ice grains from an ocean below the surface of Enceladus, Nature **459**, 1098 (2009).

[S42] F. Postberg, J. Schmidt, J. Hillier, S. Kempf, and R. Srama, A salt-water reservoir as the source of a compositionally stratified plume on Enceladus., Nature **474**, 620 (2011).

[S43] D. P. Glavin, A. S. Burton, J. E. Elsila, J. C. Aponte, and J. P. Dworkin, The Search for Chiral Asymmetry as a Potential Biosignature in our Solar System, Chemical Reviews **120**, 4660 (2020).

[S44] I. Guyon and A. Elisseeff, An introduction to variable and feature selection., Journal of Machine Learning Research **3**, 1157 (2003).

[S45] A. E. Hoerl and R. W. Kennard, Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics **12**, 55 (1970).

[S46] R. R. Hocking, A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression, Biometrics **32**, 1 (1976), publisher: [Wiley, International Biometric Society].
[S47] F. Santosa and W. W. Symes, Linear Inversion of Band-Limited Reflection Seismograms, SIAM Journal on Scientific and Statistical Computing **7**, 1307 (1986), publisher: Society for Industrial and Applied Mathematics.
[S48] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society. Series B (Methodological) **58**, 267 (1996), publisher: [Royal Statistical Society, Wiley].
[S49] S. L. Davies, A. A. Neath, and J. E. Cavanaugh, Estimation Optimality of Corrected AIC and Modified Cp in Linear Regression, International Statistical Review / Revue Internationale de Statistique **74**, 161 (2006), publisher: [Wiley, International Statistical Institute (ISI)].
[S50] M. Stone, An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion, Journal of the Royal Statistical Society: Series B (Methodological) **39**, 44 (1977).
[S51] P. Stoica, P. Eykhoff, P. Janssen, and T. Söderström, Model-structure selection by cross-validation, International Journal of Control **43**, 1841 (1986).
[S52] J. H. Waite, H. Niemann, R. V. Yelle, W. T. Kasprzak, T. E. Cravens, J. G. Luhmann, R. L. McNutt, W.-H. Ip, D. Gell, V. De La Haye, I. Müller-Wordag, B. Magee, N. Borggren, S. Ledvina, G. Fletcher, E. Walter, R. Miller, S. Scherer, R. Thorpe, J. Xu, B. Block, and K. Arnett, Ion Neutral Mass Spectrometer Results from the First Flyby of Titan, Science **308**, 982 (2005).
[S53] J. H. Waite, D. T. Young, T. E. Cravens, A. J. Coates, F. J. Crary, B. Magee, and J. Westlake, The Process of Tholin Formation in Titan's Upper Atmosphere, Science **316**, 870 (2007).
[S54] J. H. Waite, R. S. Perryman, M. E. Perry, K. E. Miller, J. Bell, T. E. Cravens, C. R. Glein, J. Grimes, M. Hedman, J. Cuzzi, T. Brockwell, B. Teolis, L. Moore, D. G. Mitchell, A. Persoon, W. S. Kurth, J.-E. Wahlund, M. Morooka, L. Z. Hadid, S. Chocron, J. Walker, A. Nagy, R. Yelle, S. Ledvina, R. Johnson, W. Tseng, O. J. Tucker, and W.-H. Ip, Chemical interactions between Saturn's atmosphere and its rings, Science **362**, eaat2382 (2018).
[S55] W. B. McKinnon, J. H. Waite, O. Mousis, J. I. Lunine, and M. Y. Zolotov, The Mysterious Origin of Enceladus: A Compositional Perspective, in *Enceladus and the Icy Moons of Saturn* (The University of Arizona Press, 2018).