

Supplementary Information

Electrical chiral discrimination of amino acids in meteorite and desert soil extracts via single-molecule nanogap conductance

Takahito Ohshiro, Yuki Komoto, Takayuki Takaai, Takeshi Yoshida, Takashi Washio, José L. Ramírez-Colón, Hannah L. McLain, Daniel Duzdevich, Maria T. Zuber, Christopher E. Carr, Masateru Taniguchi

Table of Contents

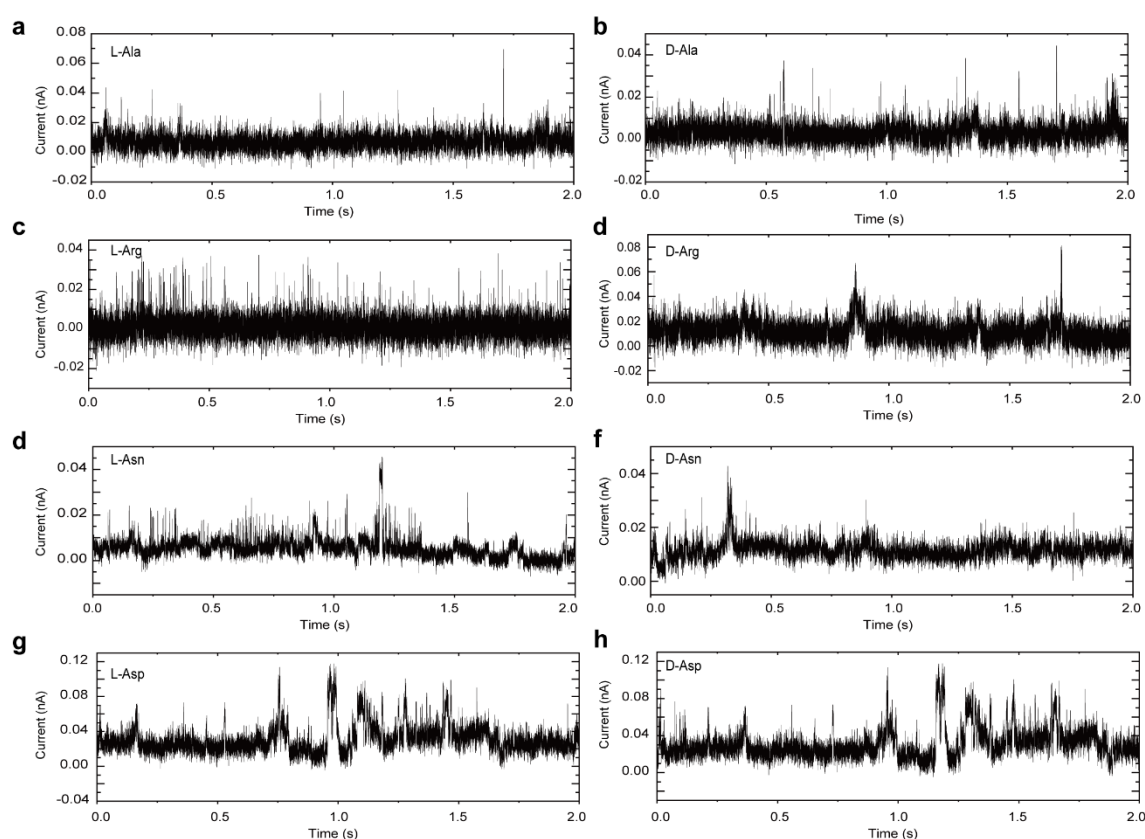
Supplementary Note 1. Signal measurement of amino acid molecules and their conductance profiles	3
Supplementary Note 2. Extraction of feature parameters from single-molecule signals	10
Supplementary Note 3. Standards and reagents for analysis of amino acids in the Murchison meteorite and Atacama Desert samples	14
Supplementary Note 4. Sample collection	14
Supplementary Note 5. Amino acid extraction	14-15
Supplementary Note 6. Single-molecule signal frequency reflects amino acid abundance in extracted environmental samples	17-18
Supplementary Note 7. Evaluation of amino acid composition in environmental samples from the Murchison meteorite, Atacama NOA, and Atacama DGA by single-molecule analysis	18-19
Supplementary Note 8. Relative quantification of amino acids in Atacama NOA and DGA samples using single-molecule detection	20
Supplementary Note 9. Chiral discrimination and D/L ratio evaluation in environmental samples from the Murchison meteorite, Atacama NOA, and Atacama DGA	21
Supplementary Note 10. LC-FD/ToF-MS analyses	22
Supplementary Note 11. Single-molecule discrimination of chiral alcohols using nanogap electrodes	24
Supplementary Note 12. Chemical and sample materials	27

Supplementary Figures and Supplementary Tables

Supplementary Figure 1. Current-time signal detection profiles of Ala, Arg, Asn, and Asp.	3
Supplementary Figure 2. Current-time signal detection profiles of Cys, Gln, Glu, and His.	4
Supplementary Figure 3. Current-time signal detection profiles of Ile, Leu, Lys, and Met.	5
Supplementary Figure 4. Current-time signal detection profiles of Phe, Pro, Ser, and Thr.	6
Supplementary Figure 5. Current-time signal detection profiles of Trp, Tyr, Val, and Gly.	7
Supplementary Figure 6. Histogram of signal time length (t_d) for L-tryptophan and D-tryptophan.	8
Supplementary Figure 7. Correlation between molecular properties, signal frequency, and t_d .	9
Supplementary Figure 8. Scheme of machine learning analysis for chiral and amino acid discrimination.	10
Supplementary Figure 9. Discrimination accuracy for chiral molecules calculated using hold-out method.	11
Supplementary Figure 10. Quantitative analysis of mixed solutions.	12
Supplementary Figure 11. Identification of chiral molecules using reused nanogap electrodes.	12-13
Supplementary Figure 12. Time dependence of displacement between nanogap electrodes.	13
Supplementary Figure 13. Summary of amino acids extraction procedure.	15
Supplementary Figure 14. Sample preparation and extraction workflow.	15
Supplementary Figure 15. Washes and dry-down workflow.	16
Supplementary Figure 16. Acid-hydrolysis workflow.	16
Supplementary Figure 17. Amino acid purification workflow.	17
Supplementary Figure 18. Sample preparation for LC-MS and nanogap analysis.	17
Supplementary Figure 19. LC-MS-ToF chromatograms of Asp, Glu, Ser, and Ala.	23
Supplementary Figure 20. Current-time profiles of four alcohols.	25
Supplementary Figure 21. Chiral discrimination of alcohols.	25
Supplementary Table 1. Classifier dependence of L-His and D-His discrimination accuracy.	11
Supplementary Table 2. Sample concentration and detected signal counts for amino acid extracts.	24
Supplementary Table 3. Event flow for confidence-filtered nanogap analysis of environmental extracts and the process blank.	27
Supplementary Table 4. Class distribution of retained high-confidence target assignments in the process blank.	27
Supplementary References	28

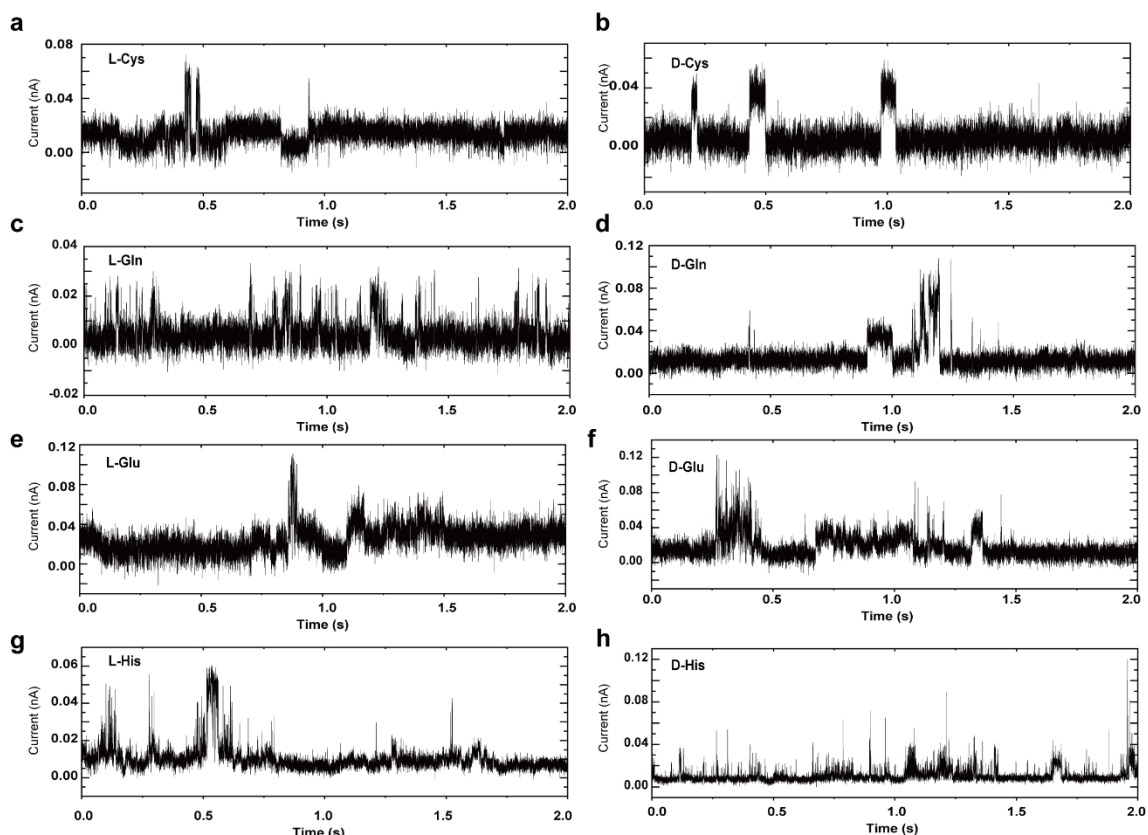
Supplementary Note 1. Signal measurement of amino acids and their conductance profiles

We measured molecular electrical signals for 39 amino acids: L-, D-alanine (L-, D-Ala), L-, D-arginine (L-, D-Arg), L-, D-asparagine (L-, D-Asn), L-, D-aspartic acid (L-, D-Asp), L-, D-cysteine (L-, D-Cys), L-, D-glutamine (L-, D-Gln), L-, D-glutamate (L-, D-Glu), L-, D-histidine (L-, D-His), L-, D-isoleucine (L-, D-Ile), L-, D-leucine (L-, D-Leu), L-, D-lysine (L-, D-Lys), L-, D-methionine (L-, D-Met), L-, D-phenylalanine (L-, D-Phe), L-, D-proline (L-, D-Pro), L-, D-serine (L-, D-Ser), L-, D-threonine (L-, D-Thr), L-, D-tryptophan (L-, D-Trp), L-, D-tyrosine (L-, D-Tyr), L-, D-valine (L-, D-Val), and glycine (Gly). The signals were measured using a nanofabricated mechanically controllable break junction (nano-MCBJ) device. The sample concentration of these aqueous amino acids was set to 0.10 μM . The actual gap size was set to about 0.54, 0.56, and 0.58 nm, and the gap distance was maintained by feedback control of the piezo actuators during all the measurements. As a result, we obtained the following conductance-time profiles (Supplementary Fig. 1–5).

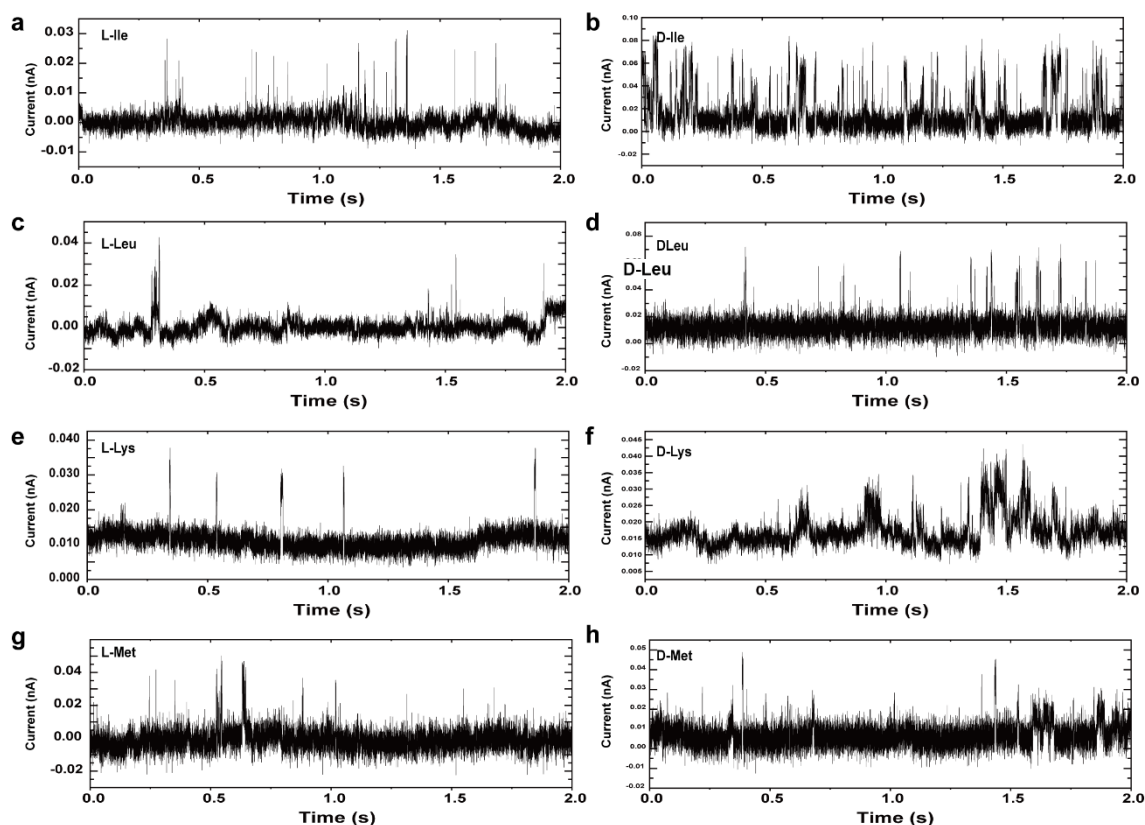


Supplementary Figure 1. Current–time signal detection profiles of alanine, arginine, asparagine, and aspartic acid. Representative current–time profiles obtained from single-molecule electrical measurements of L-alanine (L-Ala, **a**), D-alanine (D-Ala, **b**), L-arginine (L-Arg, **c**), D-arginine (D-Arg, **d**), L-asparagine (L-Asn, **e**), D-asparagine (D-Asn, **f**), L-aspartic acid (L-Asp, **g**), and D-aspartic acid (D-Asp, **h**). The nanogap distance was approximately 0.54 nm for all measurements. Representative profiles were obtained from independent measurements performed under identical experimental

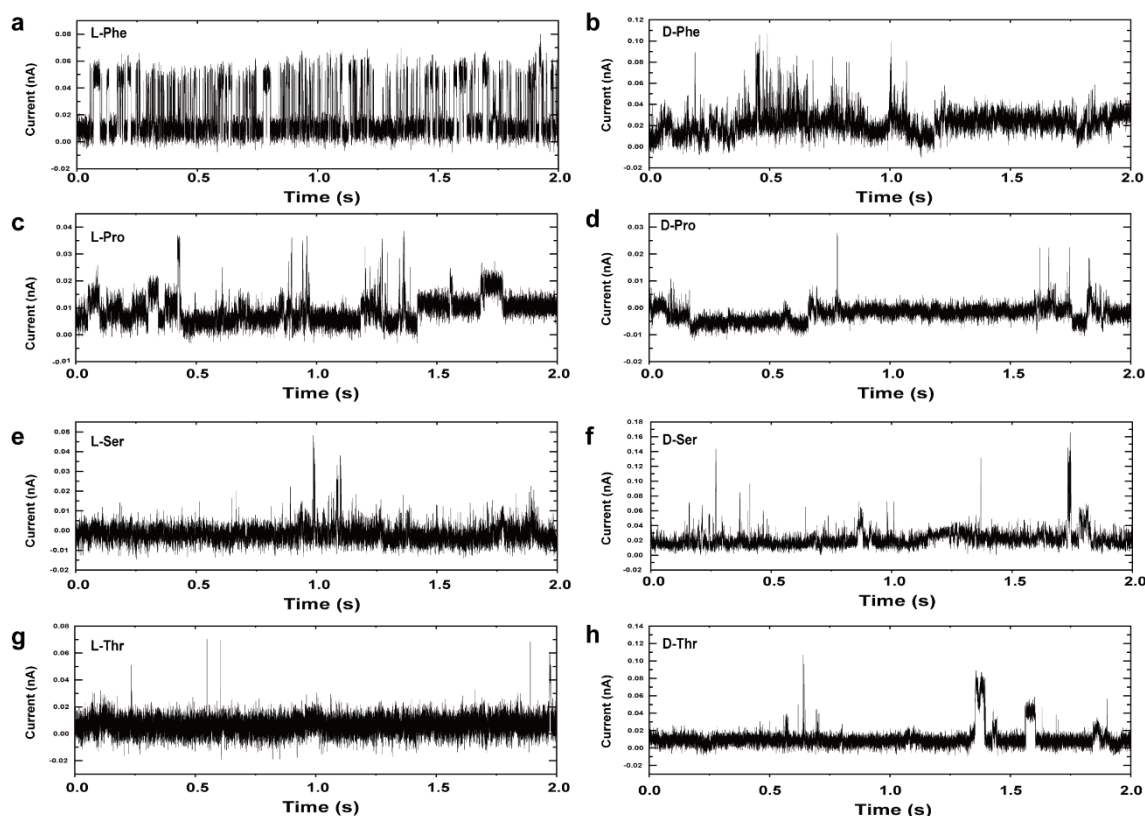
conditions. Each measurement was independently repeated $n = 3$ times under identical experimental conditions, with similar results.



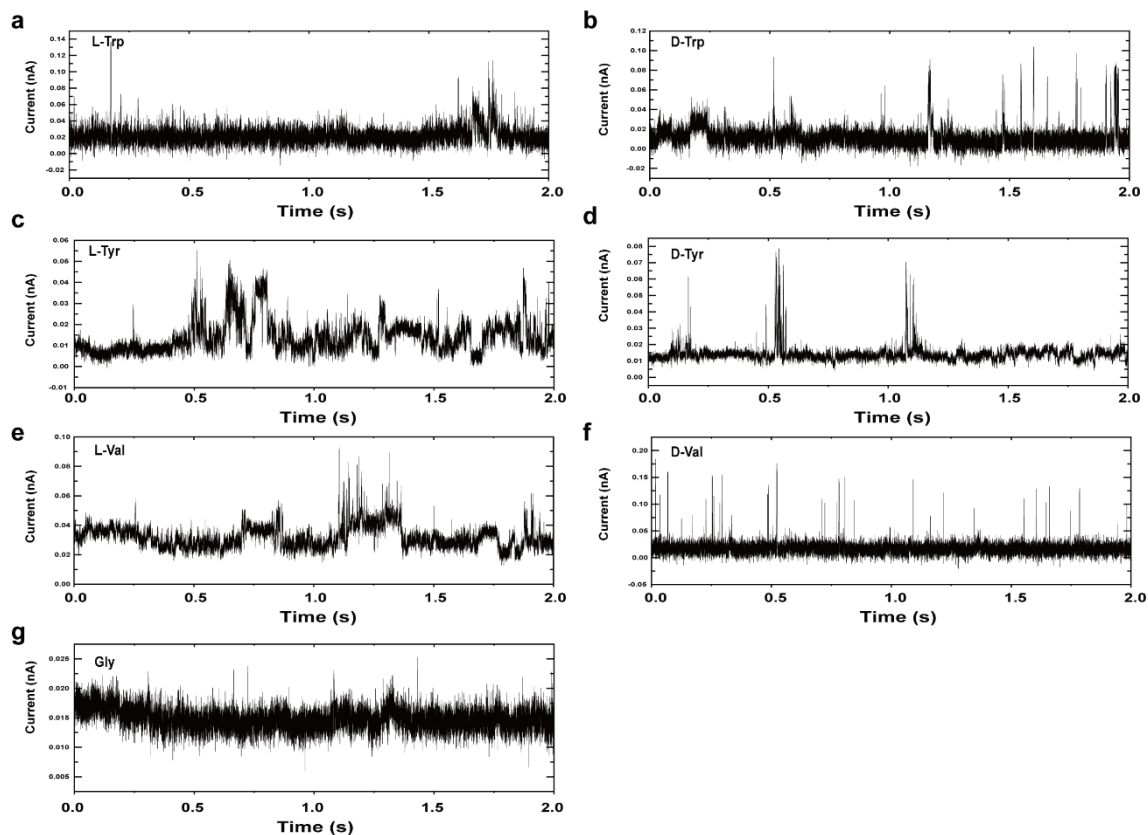
Supplementary Figure 2. Current–time signal detection profiles of Cys, Gln, Glu, and His. Single-molecule electrical signals for L-cysteine (L-Cys: a), D-cysteine (D-Cys: b), L- glutamic acid (L-Gln: c), D-glutamine (D-Gln: d), L-glutamate (L-Glu: e), D- glutamate (D-Glu: f), L-histidine (L-His: g), and D-histidine (D-His: h). The nanogap size used for these experiments was about 0.54 nm. Each measurement was independently repeated $n = 3$ times under identical experimental conditions, with similar results.



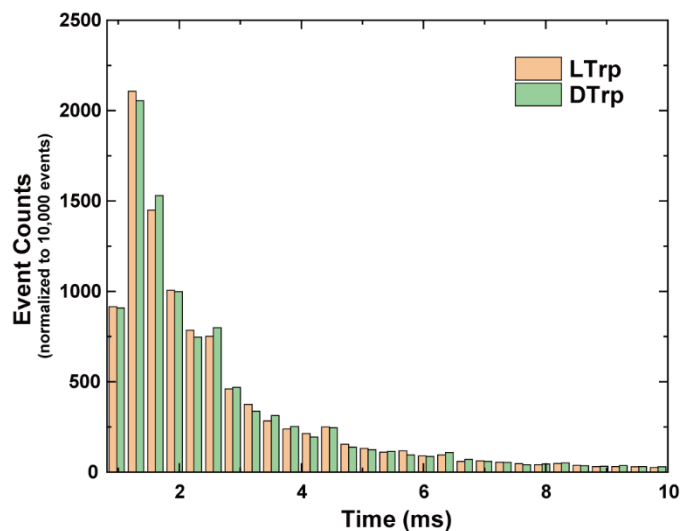
Supplementary Figure 3. Current–time signal detection profiles of isoleucine, leucine, lysine, and methionine. Representative current–time profiles obtained from single-molecule electrical measurements of L-isoleucine (L-Ile, **a**), D-isoleucine (D-Ile, **b**), L-leucine (L-Leu, **c**), D-leucine (D-Leu, **d**), L-lysine (L-Lys, **e**), D-lysine (D-Lys, **f**), L-methionine (L-Met, **g**), and D-methionine (D-Met, **h**). The nanogap distance was approximately 0.54 nm for all measurements. Each measurement was independently repeated $n = 3$ times under identical experimental conditions, with similar results.



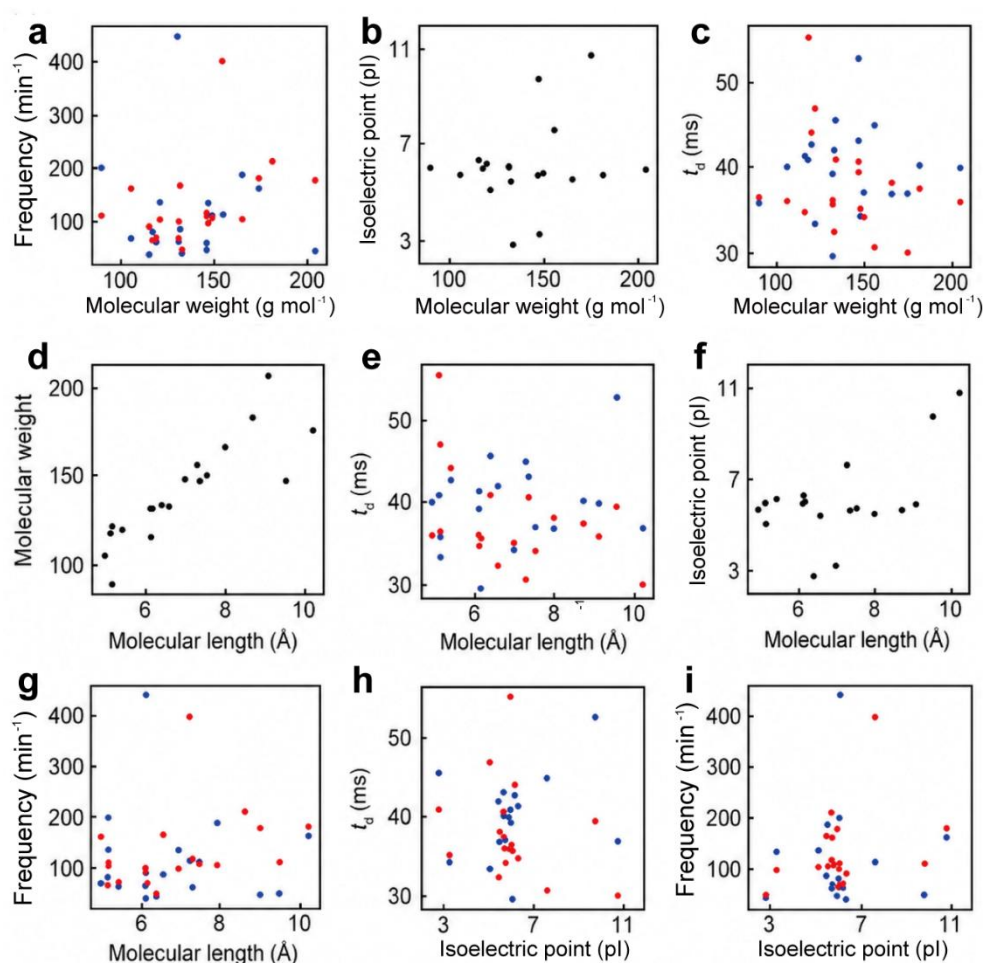
Supplementary Figure 4. Current–time signal detection profiles of phenylalanine, proline, serine, and threonine. Representative current–time profiles obtained from single-molecule electrical measurements of L-phenylalanine (L-Phe, **a**), D-phenylalanine (D-Phe, **b**), L-proline (L-Pro, **c**), D-proline (D-Pro, **d**), L-serine (L-Ser, **e**), D-serine (D-Ser, **f**), L-threonine (L-Thr, **g**), and D-threonine (D-Thr, **h**). The nanogap distance was approximately 0.54 nm for all measurements. Each experiment was independently repeated three times ($n = 3$ independent measurements) under identical experimental conditions, with similar results.



Supplementary Figure 5. Current–time signal detection profiles of tryptophan, tyrosine, valine, and glycine. Representative current–time profiles obtained from single-molecule electrical measurements of L-tryptophan (L-Trp, **a**), D-tryptophan (D-Trp, **b**), L-tyrosine (L-Tyr, **c**), D-tyrosine (D-Tyr, **d**), L-valine (L-Val, **e**), D-valine (D-Val, **f**), and glycine (Gly, **g**). The nanogap distance was approximately 0.54 nm for all measurements. Each experiment was independently repeated three times ($n = 3$ independent measurements) under identical experimental conditions, with similar results.



Supplementary Figure 6. Normalized histograms of signal duration (t_a) for L-tryptophan and D-tryptophan. To enable direct comparison of the signal-duration distributions, each histogram was normalized to a total of 10,000 tunnelling events. The original datasets contained 28,320 and 11,005 tunnelling events for L-tryptophan and D-tryptophan, respectively. The two normalized distributions exhibit similar overall profiles and peak positions, indicating comparable signal-duration distributions under the present measurement conditions.

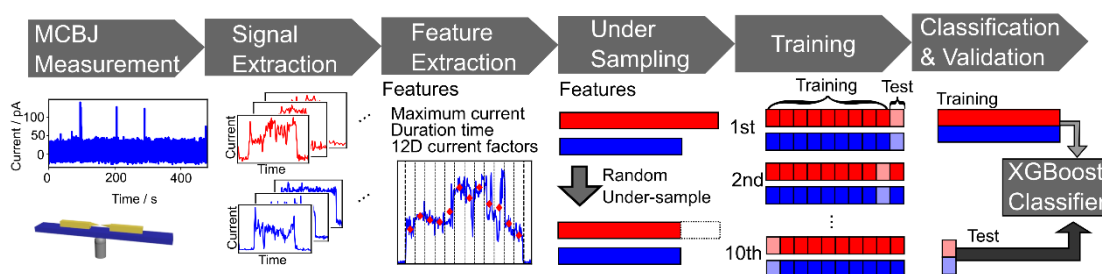


Supplementary Figure 7. Correlation between molecular properties, signal frequency, and t_d . Correlation between molecular weight and **a**, signal frequency, **b**, pI, and **c**, signal duration (t_d). Correlation between molecular length and **d**, molecular weight, **e**, signal duration (t_d), **f**, pI, and **g**, signal frequency. Correlation between pI and **h**, signal duration (t_d) and **i**, signal frequency. Pearson correlation coefficients (r) were 0.19, 0.22, -0.13 , 0.86 , -0.12 , 0.54 , 0.15 , -0.013 , and 0.18 , respectively. Each point represents one amino acid. Blue and red points indicate the L- and D-enantiomers, respectively.

Supplementary Note 2. Extraction of feature parameters from single-molecule signals

Following event detection from the current–time profiles as described in the Methods, 15 signal features were extracted from each detected event for machine-learning analysis. These comprised the peak current amplitude relative to the baseline (I_p), mean current amplitude (I_m), event duration (t_d), and 12 normalized shape features (T_1 – T_{12}). To obtain the shape features, each event was divided into 12 equal-duration intervals along the time axis. The mean current amplitude within each interval was calculated and normalized by the peak current amplitude I_p , yielding the 12 dimensionless shape parameters $T_1, T_2, \dots, T_{12}$. Thus, each event was represented by the 15-dimensional feature vector ($I_p, I_m, t_d, T_1, T_2, \dots, T_{12}$).

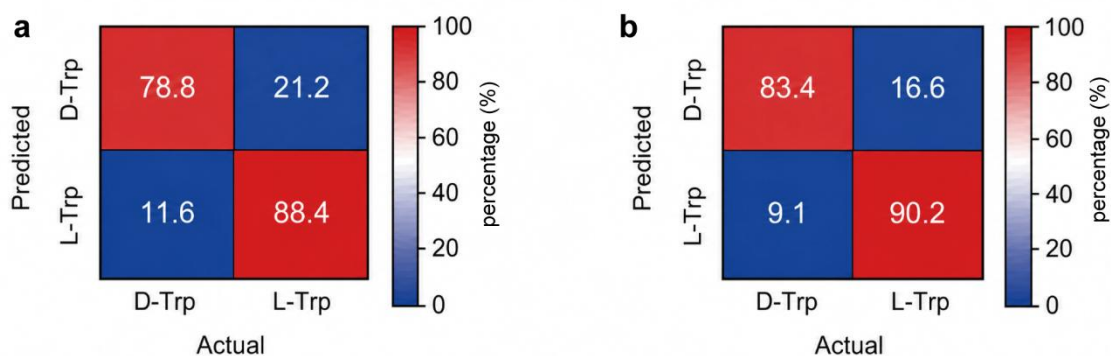
Machine-learning classification was performed using the XGBoost classifier in Python 3.9.7 with scikit-learn version 0.24.2, as described in the Methods. After random undersampling to reduce class imbalance, the balanced dataset was randomly divided into training (80%) and held-out test (20%) datasets. Ten-fold cross-validation was performed within the training dataset, with nine folds used for model training and the remaining fold used for validation in each round. Classification performance was evaluated on the held-out test dataset using the F_1 -score and confusion matrix.



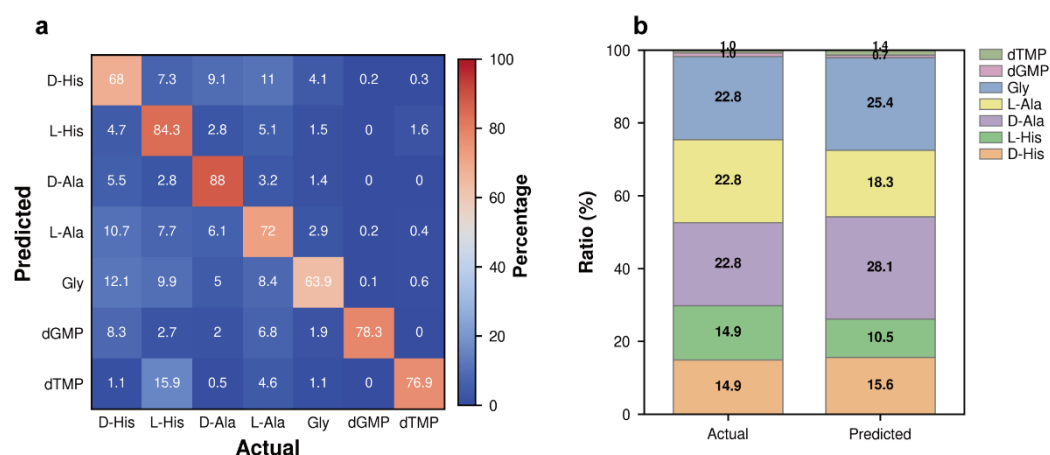
Supplementary Figure 8. Workflow of machine-learning analysis. Single-molecule electrical signals were acquired using a mechanically controllable break junction (MCBJ) device. Following signal extraction, features were generated by temporal segmentation and signal normalization. The dataset was balanced by random undersampling and divided into training (80%) and held-out test (20%) datasets. Model training and validation were performed using an Extreme Gradient Boosting (XGBoost) classifier with 10-fold cross-validation within the training dataset, and classification performance was evaluated using the held-out test dataset.

Supplementary Table 1. Classifier dependence of L-Trp and D-Trp discrimination performance. L-Trp and D-Trp were measured using 0.54 nm nanogap electrodes. 12540 and 23877 events were obtained, respectively. Machine learning was performed according to the scheme shown in Supplementary Fig. 8.

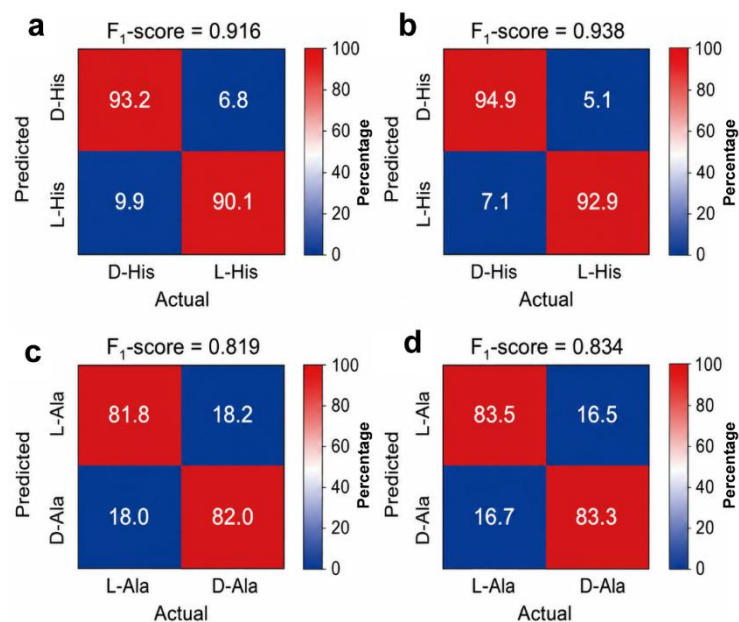
Classifier	F ₁ -score
XGBoost	0.876
Random Forest	0.813
Gradient Boosting	0.807
SVM	0.789
Neural Network	0.774
AdaBoost	0.771
Logistic Regression	0.757
Decision Tree	0.751
Naive Bayes	0.734
K-Nearest Neighbors	0.683



Supplementary Figure 9. Discrimination accuracy for chiral molecules calculated using hold-out method. The training and test datasets were obtained using different nanogap chips. **a**, Confusion matrix obtained during the training process. **b**, Confusion matrix obtained during the test process. The values in each confusion matrix represent the percentage (%) of current-time waveforms assigned to each class. The color scale also represents the percentage (%) of waveforms assigned to each class.

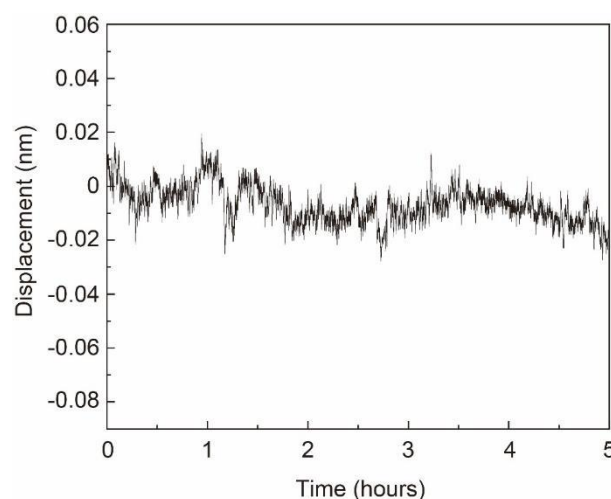


Supplementary Figure 10. Quantitative analysis of mixed solutions containing amino acids and nucleotides. **a**, Confusion matrix for a mixed solution containing D-histidine (D-His), L-histidine (L-His), D-alanine (D-Ala), L-alanine (L-Ala), glycine (Gly), deoxyguanosine monophosphate (dGMP), and deoxythymidine monophosphate (dTMP). Values indicate the percentage of classified current-time waveforms assigned to each class. The color scale represents the percentage of waveforms assigned to each class. **b**, Quantitative analysis of the mixed solution containing D-His, L-His, D-Ala, L-Ala, Gly, dGMP, and dTMP. The stacked bars show the relative composition (%) of the prepared sample (Actual) and the composition estimated from the classified current-time waveforms (Predicted).



Supplementary Figure 11. Identification of chiral molecules using reused nanogap electrodes. Chiral discrimination of the L- and D-forms of alanine (Ala) and histidine (His) was performed using the same nanogap device. After the first measurement, the nanogap electrodes were rinsed with ultrapure water before a second measurement was performed using the same device. Machine-learning models were trained independently using the datasets obtained from the first and second measurements. Confusion matrices

for histidine are shown in **a** (first measurement; F_1 -score = 0.916) and **b** (second measurement; F_1 -score = 0.938), and those for alanine are shown in **c** (first measurement; F_1 -score = 0.819) and **d** (second measurement; F_1 -score = 0.834). Comparable discrimination performance was obtained after reuse of the nanogap electrodes. The values in each confusion matrix represent the percentage (%) of current–time waveforms assigned to each class. The color scale also represents the percentage (%) of waveforms assigned to each class.



Supplementary Figure 12. Time dependence of displacement between nanogap electrodes. The plot shows the displacement required to maintain a constant nanogap distance during a continuous 5-h measurement. The displacement remained within approximately ± 0.02 nm of the initial position throughout the experiment, demonstrating stable mechanical positioning of the nanogap electrodes.

Supplementary Note 3. Standards and reagents for analysis of amino acids in the Murchison meteorite and the Atacama Desert samples.

All glassware, ceramics, and sample handling tools were rinsed with Milli-Q ultrapure water ($18.2 \text{ M}\Omega \cdot \text{cm}$, $< 3 \text{ ppb}$ total organic carbon), wrapped in aluminum foil, and heated in a furnace at 500°C in air overnight. Most of the chemicals and reagents were purchased from Sigma-Aldrich. A stock amino acid solution was prepared by mixing individual amino acid standards (97–99 % purity) in Milli-Q ultrapure water. All chiral amino acid standards were purchased as racemic mixtures (where the D-amino acid concentration is equal to the L-amino acid concentration). Acid vapor hydrolysis used Tamapure-AA-10-HCl 20 %, a high-purity hydrochloric acid reagent (metallic impurity $< 10 \text{ pg mL}^{-1}$) to minimize metal contamination during amino acid hydrolysis. Cation-exchange resin (AG50W-X8, 100–200 mesh, NH_4OH) was used. The 2 M NaOH was prepared by dissolving 32 g of NaOH pellets (Sigma-Aldrich) in 400 mL Milli-Q ultrapure water, the 1.5 N HCl was prepared from diluting the Tamapure HCl, the 2 M NH_4OH was prepared in vacuo using Milli-Q ultrapure water and ammonia gas (Air Products).

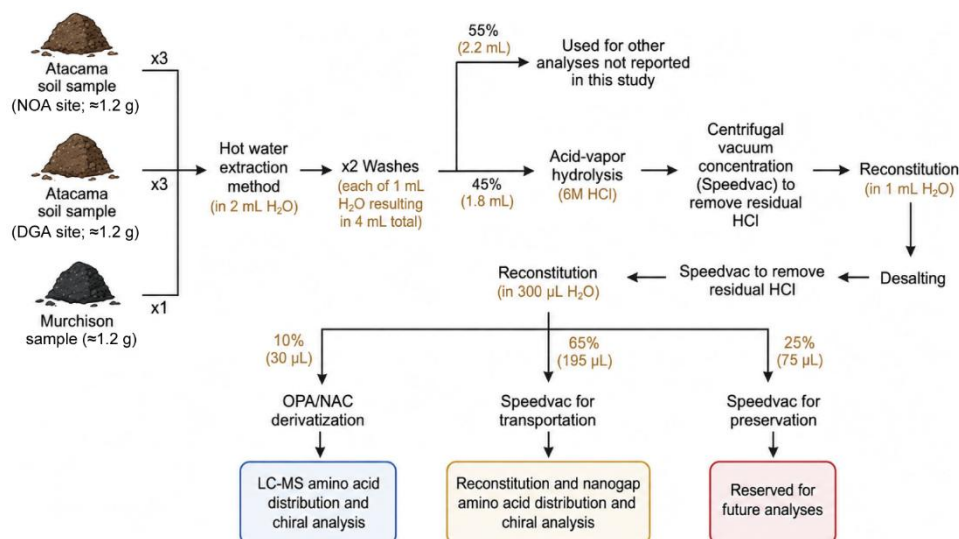
Supplementary Note 4. Samples collection.

Atacama Desert samples were collected during a field expedition in February 2022. Two sampling sites were targeted: Diego de Almagro (DGA; $26^\circ 23' 11'' \text{ S}$, $70^\circ 04' 14'' \text{ W}$) and North of Antofagasta (NOA; $22^\circ 47' 46'' \text{ S}$, $70^\circ 04' 14'' \text{ W}$). At each site, material was collected from depths of 5–10 cm below the surface, transferred into sterile Whirl-Pak® bags, and stored at ambient temperature until further processing. In addition, a 2 g interior fragment of the Murchison meteorite, free of fusion crust, was obtained from the Smithsonian National Museum of Natural History for use in extractions.

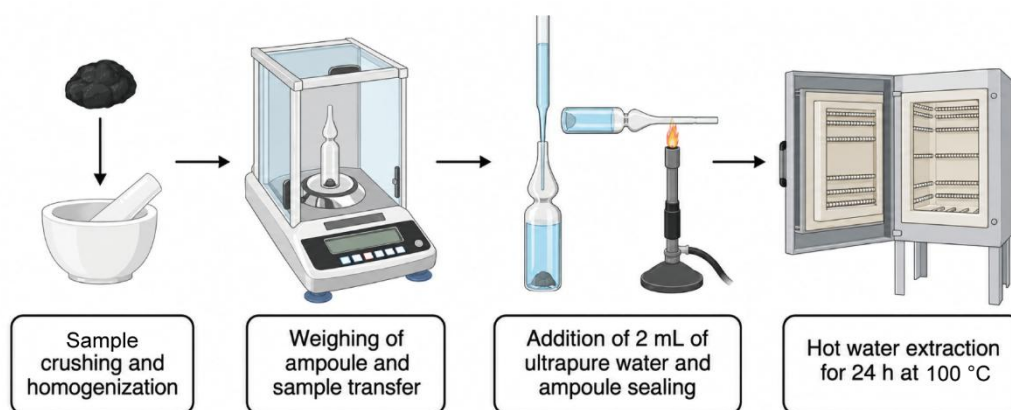
Supplementary Note 5. Amino acids extraction.

The Murchison meteorite (1.2174 g), Atacama DGA sample a (1.1943 g), Atacama DGA sample b (1.2109 g), Atacama DGA sample c (1.1867 g), Atacama NOA sample a (1.2336 g), Atacama NOA sample b (1.2033 g), Atacama NOA sample c (1.2428 g), Fused silica blank (1.2190 g), and procedural blank samples were flame-sealed in glass ampoules in 1 mL of Milli-Q ultrapure water and then heated at 100°C for 24 h (Supplementary Fig. 14). After heating, the ampoules were opened and centrifuged at $3000 \times g$ for three minutes, the water supernatants were then transferred from the ampoules by pipetting into pre-weighed amber glass vials (Supplementary Figure 15). Another 1 mL of Milli-Q ultrapure water was added to each glass ampoule, the ampoules were re-centrifuged, and the supernatant was transferred to the sample amber glass vials (this process was repeated one final time to maximize the recovery of the water extracts). After extraction, 45 % of the supernatant was dried down under vacuum and subsequently subjected to 6 M HCl vapor hydrolysis at 150°C for 3 h (Supplementary Fig. 16). The HCl acid-hydrolyzed and 45 % of the non-acid vaporized supernatant were desalted using AG50W-X*, 100–200 mesh, hydrogen form, BIO-RAD cation exchange resin. Both the vapor hydrolyzed (45 %) and non-vapor hydrolyzed (45 %) supernatant sample sets were eluted from the cation exchange desalting columns using 7 mL of 2 M NH_4OH (Supplementary Fig. 17). After desalting, the eluents were dried under vacuum and resuspended in 100 μL of Milli-Q ultrapure water separately (100 μL non-hydrolyzed and 100 μL hydrolyzed). The samples were then vortexed and transferred to vials. Two washes of 200 μL of Milli-Q ultrapure water were performed on each of the dried-down

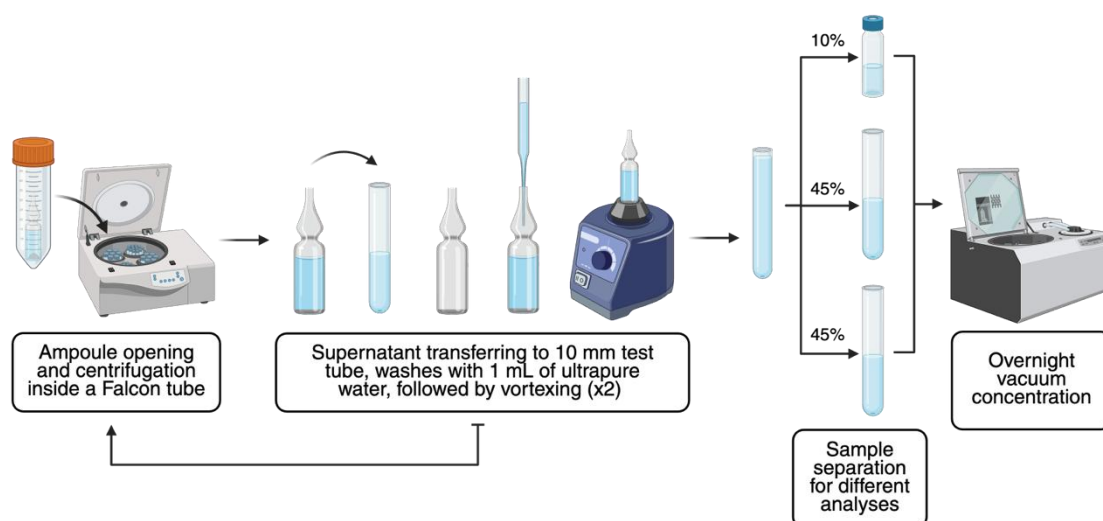
samples tubes to maximize the recovery of the dried-down desalted extracts. The resulting volumes of 300 μL for each analysis (non-hydrolyzed and hydrolyzed) were then divided into three parts: 10% for liquid chromatography-mass spectrometry (LC-MS) analysis, 65% for single-molecule nanogap analysis, and 25% for additional tests not reported in this study (Supplementary Figure 18).



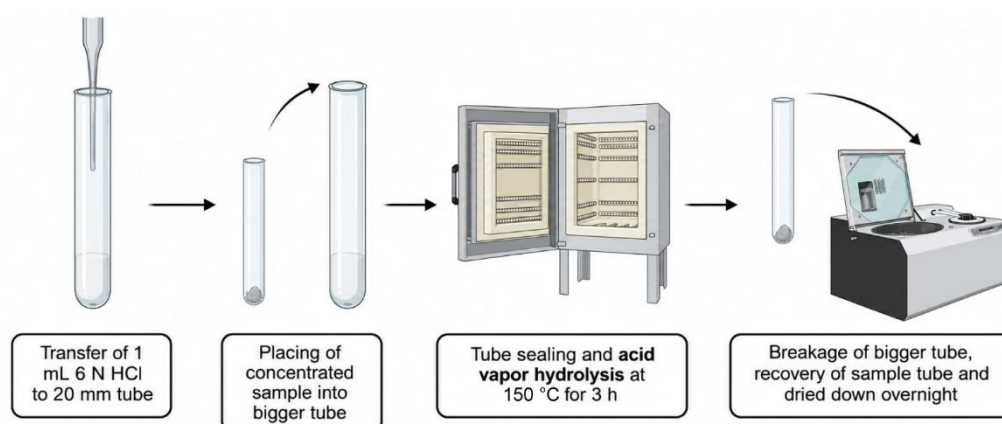
Supplementary Figure 13. Summary of amino acids extraction procedure. Schematic overview of the extraction workflow applied to the Murchison meteorite fragment and the two Atacama Desert soil samples (NOA and DGA). The workflow includes hot-water extraction, washing, acid-vapor hydrolysis, desalting, reconstitution, and aliquoting. Volumes and percentages allocated for LC-MS analysis, nanogap measurements, and future analyses are indicated.



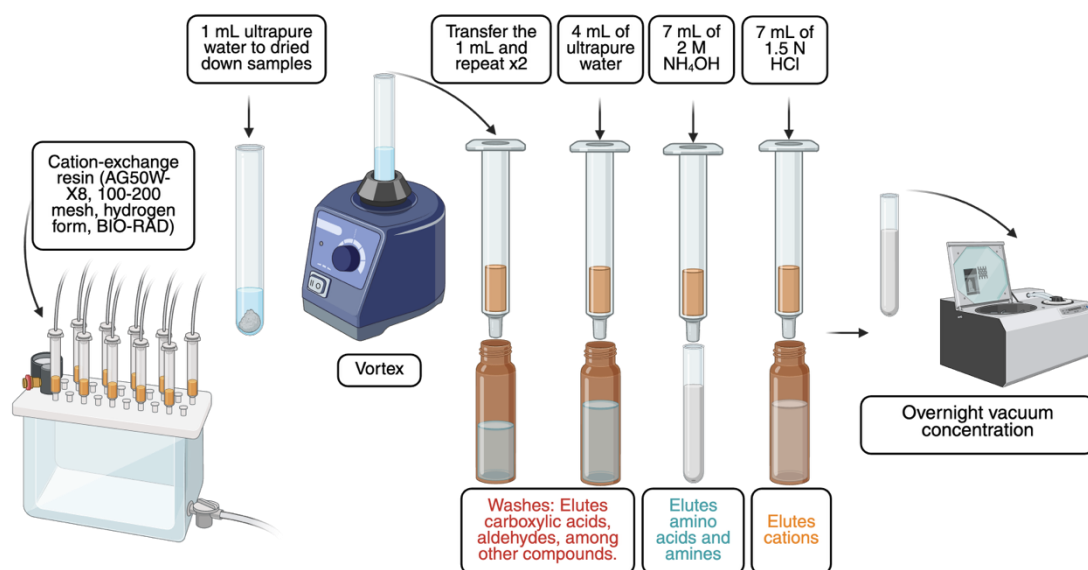
Supplementary Figure 14. Sample preparation and extraction workflow. Schematic illustration of the sample preparation and hot-water extraction procedure for the Murchison meteorite and Atacama Desert soil samples. The workflow includes sample crushing, suspension in ultrapure water, hot-water extraction, centrifugation, and collection of the supernatant for subsequent analyses.



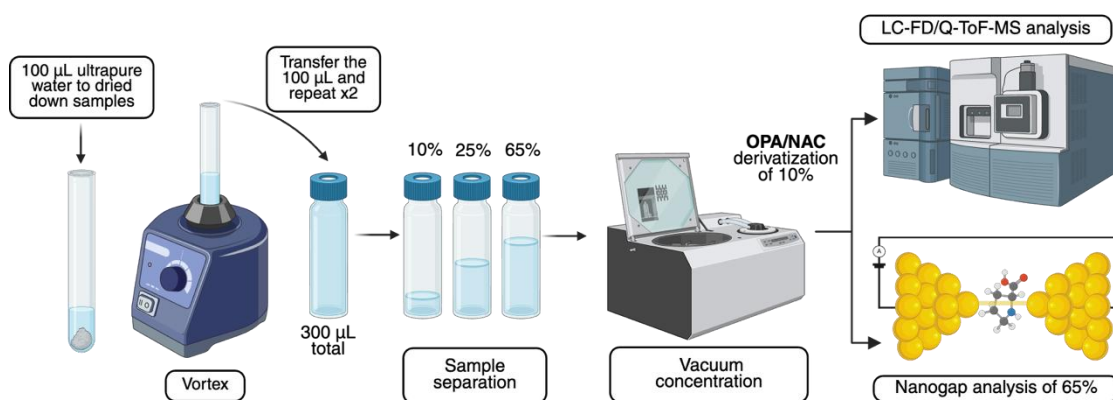
Supplementary Figure 15. Washing and dry-down workflow. Schematic illustration of the washing, sample-splitting, and dry-down procedure. After ampoule opening and centrifugation, the supernatant was transferred to a 10-mL test tube and the original tube was washed twice with 1 mL of ultrapure water using vortex mixing. The combined solution was then divided into aliquots for non-hydrolyzed analysis (45%), hydrolyzed analysis (45%), and additional analyses (10%, not included in this study), followed by overnight vacuum concentration.



Supplementary Figure 16. Acid-hydrolysis workflow. Schematic illustration of the vapor-phase acid hydrolysis procedure applied to the hot-water extracts of the Murchison meteorite and Atacama Desert soil samples. Sample aliquots were sealed under vacuum with 6 M HCl and hydrolyzed at 150 °C for 3 h, followed by vacuum concentration prior to subsequent analyses.



Supplementary Figure 17. Amino acid purification workflow. Schematic illustration of the cation-exchange purification procedure applied to the environmental sample extracts. Samples were loaded onto AG50W-X8 cation-exchange resin, washed to remove carboxylic acids, aldehydes, and other neutral compounds, and sequentially eluted to obtain amino acid/amine and cation fractions. Only the amino acid/amine fraction was used for the LC-MS and nanogap analyses reported in this study.



Supplementary Figure 18. Sample preparation for LC-MS and nanogap analysis. Schematic illustration of the washing and sample preparation procedure applied to the desalted dry-down samples. After repeated washing with ultrapure water and vacuum concentration, the samples were divided into aliquots for LC-MS analysis (10%) and nanogap measurements (65%), additional analyses (25%).

Supplementary Note 6. Single-molecule signal frequency reflects amino acid abundance in extracted natural samples from the Murchison meteorite and the Atacama Desert.

To quantify amino acid concentrations in the natural samples, continuous single-molecule tunneling current measurements were conducted for 15 min per sample at

nanogap distances of 0.58, 0.56, and 0.54 nm. The detection limit of the measurement system was estimated to be approximately 1.0×10^{-11} mol/L. Signal events were defined as current spikes exceeding the baseline noise, and their frequency was used as an indicator of molecular abundance.

As an initial test, the resuspended acid-hydrolyzed fraction of the Murchison meteorite extract was analyzed. Based on LC-MS measurements (Supplementary Table 2), the concentration of the amino acids quantified in this sample was estimated at 28.1 μ M. Under these conditions, a total of 41,158 single-molecule signal events were recorded across three measurement trials (Fig. 5a). Although the solution contained approximately 2.538×10^{12} amino acid molecules, only about 1.62 microperecent ($\mu\%$)—**corresponding to molecules that diffused into and transiently occupied the nanogap volume (~femtoliter scale)**—were detected as signal events. This fraction reflects the limited sampling of molecules physically entering the tunneling junction during the measurement time, rather than incomplete ergodicity of the solution. Such limited sampling is intrinsic to single-molecule measurements of highly dilute systems.

Similar single-molecule measurements were conducted in triplicates from the Atacama NOA and DGA sites, along with a blank control. In many cases, a clear correlation was observed between the estimated amino acid concentrations and the frequency of detected signal events. However, despite their relatively low estimated concentrations, both the blank and NOA samples exhibited high signal frequencies. This discrepancy suggests that background contamination may have contributed substantially to the signal output under low-concentration conditions. Possible sources include residual organic molecules on the electrode surface, adsorption of airborne contaminants during chip handling, or trace impurities in the solvent. Although higher concentrations or smaller sample volumes could, in principle, improve the signal-to-noise ratio, a minimum volume (~ 10 μ L) was required to maintain stable electrical contact and prevent electrode drying during long-term measurements.

Supplementary Note 7. Confidence-filtered evaluation of amino acid composition in natural samples from the Murchison meteorite and Atacama Desert

The amino acid compositions of extracts from the Murchison meteorite and the North of Antofagasta (NOA) and Diego de Almagro (DGA) sites in the Atacama Desert were evaluated by single-molecule nanogap measurements and LC-MS analysis. A previous study by Amashukeli et al.³ identified glutamate, aspartic acid, serine, valine, alanine, proline and glycine in Atacama Desert soil samples. Based on this precedent and compatibility with the LC-MS workflow, the present study focused on glycine and the D- and L-enantiomers of alanine, valine, serine, aspartic acid and glutamate. Proline was excluded because D- and L-proline could not be quantified under the LC-MS derivatization workflow used in this study. Glycine was treated as a single achiral target, whereas the other five amino acids were treated as separate D- and L-enantiomer targets, resulting in a predefined 11-species panel.

The environmental-sample analysis was designed as targeted relative quantification within this predefined panel and was not intended to provide comprehensive profiling of all amino acids or other organic constituents in the extracts. Current-time signals were processed to obtain 15-dimensional feature vectors, and feature-extractable events were classified using the trained XGBoost model. For each event, the maximum predicted probability across the 11 target classes was calculated. Events with a maximum predicted

probability of at least 0.5 were retained as high-confidence target assignments, whereas events with a maximum probability below 0.5 were categorized as unknown/rejected and excluded from the compositional and D/L-ratio analyses.

The proportions of retained high-confidence target events among feature-extractable events were 49.7% for Murchison, 28.3% for DGA, 25.4% for NOA and 6.8% for the process blank (Supplementary Table 5). Thus, the majority of feature-extractable blank events were categorized as unknown/rejected under the same criterion. The class distribution of retained high-confidence blank assignments was non-uniform, indicating that residual process-background or non-target matrix signals can contribute unequally to individual target classes (Supplementary Table 3).

For the Murchison extract, the confidence-filtered nanogap composition was glycine, 47.9%; alanine, 16.9%; valine, 8.2%; serine, 11.7%; aspartic acid, 3.3%; and glutamate, 12.0% (Fig. 5b and Table 3). LC–MS analysis of the same extract gave relative abundances of 62.84% for glycine, 15.28% for alanine, 2.55% for valine, 0.34% for serine, 2.57% for aspartic acid and 16.43% for glutamate. Thus, the confidence-filtered nanogap analysis retained the major glycine-, alanine- and glutamate-rich features observed by LC–MS, while analyte-specific differences remained, particularly for serine and valine.

The NOA and DGA extracts also showed sample-specific high-confidence target-panel compositions, but agreement with LC–MS varied according to analyte and sample matrix. In particular, the lower high-confidence target fractions observed for NOA and DGA indicate a greater relative contribution of process-background and/or non-target matrix events than in Murchison. Therefore, the nanogap-derived compositions should be interpreted as relative compositions within the retained high-confidence target panel rather than as absolute amino acid compositions of the environmental extracts. Matrix-matched calibration, increased event statistics, representative non-target reference data and open-set recognition approaches will be required to establish practical quantitative performance in heterogeneous environmental samples.

Supplementary Note 8. Confidence-filtered relative composition analysis of amino acids in Atacama NOA and DGA samples using single-molecule detection

In addition to the Murchison extract, acid-hydrolyzed extracts from the Atacama North of Antofagasta (NOA) and Diego de Almagro (DGA) sites were analysed by nanogap single-molecule detection and LC–MS. The nanogap analysis was performed as targeted relative quantification within the predefined 11-species panel comprising glycine and the D- and L-enantiomers of alanine, valine, serine, aspartic acid and glutamate. For each feature-extractable event, the maximum predicted probability across the 11 target classes was calculated. Events with a maximum probability of at least 0.5 were retained as high-confidence target assignments, whereas events below this threshold were categorized as unknown/rejected and excluded from the compositional analysis.

The retained high-confidence target fraction was 25.4% for NOA and 28.3% for DGA, indicating that a substantial proportion of feature-extractable events did not meet the confidence criterion for assignment to the targeted amino-acid panel. Accordingly, the nanogap-derived values represent relative compositions within the retained high-confidence target events and should not be interpreted as absolute amino-acid compositions of the environmental extracts.

For NOA, the confidence-filtered nanogap composition was glycine, 56.4%; alanine, 7.5%; valine, 13.5%; serine, 10.3%; aspartic acid, 12.2%; and glutamate, 0.0% (Fig. 5b and Table 3). LC–MS analysis showed glycine as the dominant component and indicated lower relative valine abundance than the nanogap analysis. Alanine, serine and aspartic acid showed broadly comparable relative contributions between the two methods, whereas glutamate was not retained as a high-confidence nanogap assignment. These results indicate analyte-dependent agreement rather than uniform quantitative correspondence.

For DGA, the confidence-filtered nanogap composition was glycine, 52.6%; alanine, 16.6%; valine, 14.5%; serine, 4.1%; aspartic acid, 12.1%; and glutamate, 0.0% (Fig. 5b and Table 3). Glycine and alanine showed similar relative contributions in the nanogap and LC–MS datasets, whereas valine, serine and aspartic acid were higher in the nanogap-derived composition. Glutamate was detected by LC–MS but was not retained as a high-confidence nanogap assignment. Thus, as observed for NOA, the level of agreement depended on individual analytes and the sample matrix.

Taken together, the confidence-filtered analyses of NOA and DGA demonstrate that the nanogap platform can retain targeted amino-acid assignments in complex environmental extracts, while also revealing substantial matrix- and analyte-dependent limitations. The lower high-confidence target fractions relative to Murchison, together with the deviations from LC–MS for selected amino acids, indicate that further matrix-matched calibration, increased event statistics, representative non-target reference data and open-set recognition approaches will be necessary before practical quantitative performance can be established for heterogeneous environmental samples.

Supplementary Note 9. Confidence-filtered chiral analysis and D/L-ratio evaluation in natural samples from the Murchison meteorite and Atacama Desert sites

Chiral analysis of amino acids in extracts from the Murchison meteorite and the North of Antofagasta (NOA) and Diego de Almagro (DGA) sites in the Atacama Desert was performed using LC–MS and nanogap single-molecule detection. The targeted nanogap analysis included the D- and L-enantiomers of alanine, valine, serine, aspartic acid and glutamate. For the nanogap analysis, D/L ratios were calculated only from retained high-confidence D- and L-enantiomer assignments within the predefined 11-species target panel. Events with a maximum predicted probability below 0.5 were categorized as unknown/rejected and excluded from the D/L-ratio analysis.

After confidence filtering, directly comparable D/L ratios were available for alanine, valine and aspartic acid in the Murchison and DGA extracts (Fig. 5c and Table 4). Serine was not included because D/L ratios could not be robustly determined under the confidence-filtered analysis. Glutamate was not included in Fig. 5c because a corresponding confidence-filtered nanogap D/L ratio was not available for DGA. NOA was not included in the direct comparison because comparable LC–MS D/L ratios were not available for the relevant analytes.

For the Murchison extract, the LC–MS and confidence-filtered nanogap D/L ratios were 1.66 ± 0.15 and 0.32 ± 0.37 for alanine, 0.35 ± 0.35 and 1.44 ± 0.27 for valine, and 0.91 ± 0.04 and 0.71 ± 0.41 for aspartic acid, respectively. Thus, the nanogap-derived ratio for aspartic acid was similar to the LC–MS value, whereas larger differences remained for alanine and valine.

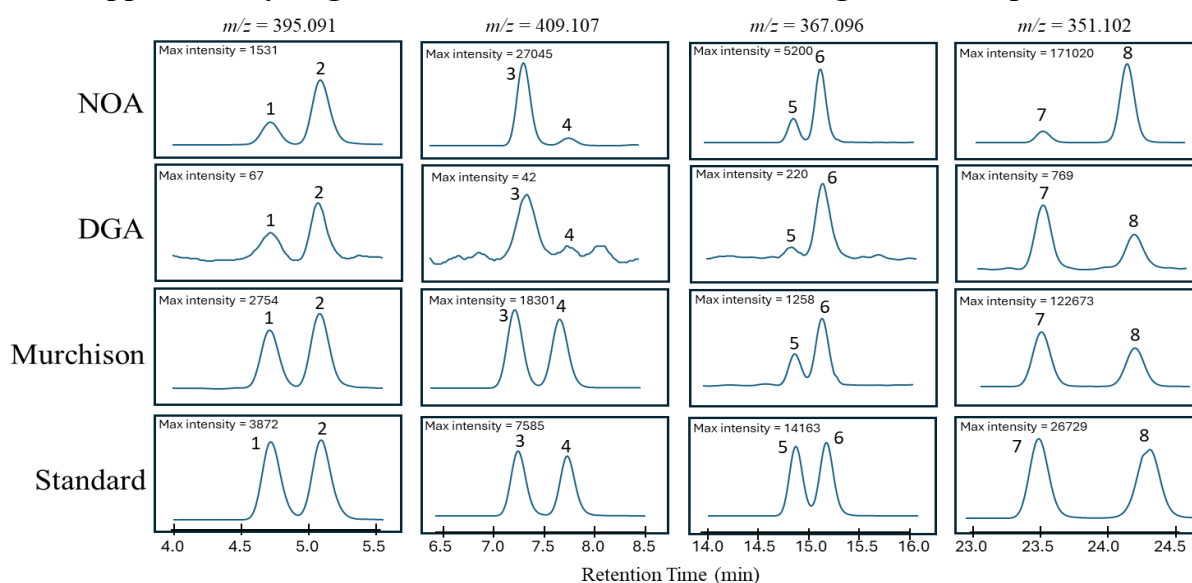
For the DGA extract, the LC–MS and confidence-filtered nanogap D/L ratios were 0.23 ± 0.12 and 0.42 ± 0.46 for alanine, 0.02 ± 0.01 and 0.34 ± 0.31 for valine, and 0.31 ± 0.12 and 0.01 ± 0.46 for aspartic acid, respectively. These comparisons showed analyte-dependent deviations between the two methods. In particular, the relatively large uncertainties associated with several nanogap-derived ratios indicate that the retained event statistics and classification uncertainty remain limiting factors for quantitative chiral analysis in these complex extracts.

Accordingly, the confidence-filtered nanogap D/L ratios should be interpreted as targeted high-confidence estimates within the predefined amino-acid panel rather than as comprehensive chiral profiles of the environmental samples. The present data do not support a global linear calibration relationship between nanogap-derived and LC–MS D/L ratios across all analytes and samples. Additional matrix-matched validation, increased event statistics, representative non-target reference datasets and open-set recognition approaches will be required to establish practical quantitative performance for chiral analysis in heterogeneous environmental samples.

Supplementary Note 10. LC-FD/ToF-MS analyses.

Detailed LC-FD/ToF-MS experimental conditions are provided in the Methods section of the main manuscript. Representative chromatograms obtained for the environmental samples and amino acid standards are shown in Supplementary Fig. 19. The amino acid abundances and their enantiomeric ratios in the meteorite extracts and controls were determined by comparison of the peak areas generated from the sample and control UV fluorescence chromatograms (LC-FD, $\lambda_{\text{ex}} = 340$ nm, $\lambda_{\text{em}} = 450$ nm) of their OPA/NAC derivatives to the corresponding peak areas of amino acid standards run under the same chromatographic conditions and included peak identification confirmation by accurate mass using a match tolerance of 10 ppm.

Supplementary Figure 19. LC-FD/ToF-MS chromatograms of aspartic acid,



glutamic acid, serine, and alanine. Selected-ion chromatograms showing the elution of aspartic acid ($m/z = 395.0913$), glutamic acid ($m/z = 409.1069$), serine ($m/z = 367.0964$), and alanine ($m/z = 351.1015$) in acid-hydrolyzed hot-water extracts of the Murchison meteorite, Atacama DGA sample, Atacama NOA sample, and amino acid standards. Chromatograms correspond to the OPA/NAC derivatives analyzed by LC-FD/ToF-MS in positive-ion mode. Similar selected-ion chromatograms were obtained for the non-hydrolyzed hot-water extracts. Peaks were assigned by comparison of retention times and exact monoisotopic masses with amino acid standards analyzed on the same day. Peak assignments are: 1, D-aspartic acid; 2, L-aspartic acid; 3, L-glutamic acid; 4, D-glutamic acid; 5, D-serine; 6, L-serine; 7, D-alanine; and 8, L-alanine.

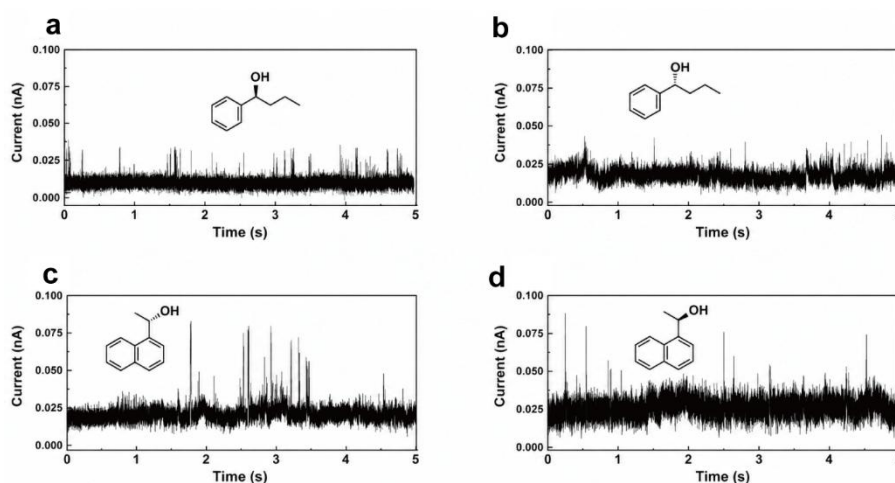
Supplementary Table 2. Sample concentration and detected signal counts for amino acid extracts. Summary of amino acid concentrations, reconstitution volumes, resulting concentrations, and the number of single-molecule tunneling signals detected for each sample. Samples include acid-hydrolyzed extracts from Murchison, Atacama NOA samples, Atacama DGA samples, and a blank control. Average values for the NOA and DGA sites are also included. Detected signal counts reflect the total number of tunneling events recorded over 15-minute measurements.

Sample name	Total [AA] (nmol)*	Reconst. Vol. (μL)	[AA] After Recons. (μM)	Signal Count (n)
Murchison	281.10	10000.00	28.10	41158
Blank	2.08	1000.00	2.08	10790
NOA a	1.63	1000.00	1.63	9058
NOA b	1.60	1000.00	1.60	6161
NOA c	2.54	1000.00	2.54	24594
DGA a	11.87	1000.00	11.87	30468
DGA b	24.42	1000.00	24.42	20027
DGA c	12.85	1000.00	12.85	19393
Averaged NOA	1.92	1000.00	1.92	13271
Averaged DGA	16.37	1000.00	16.37	23296

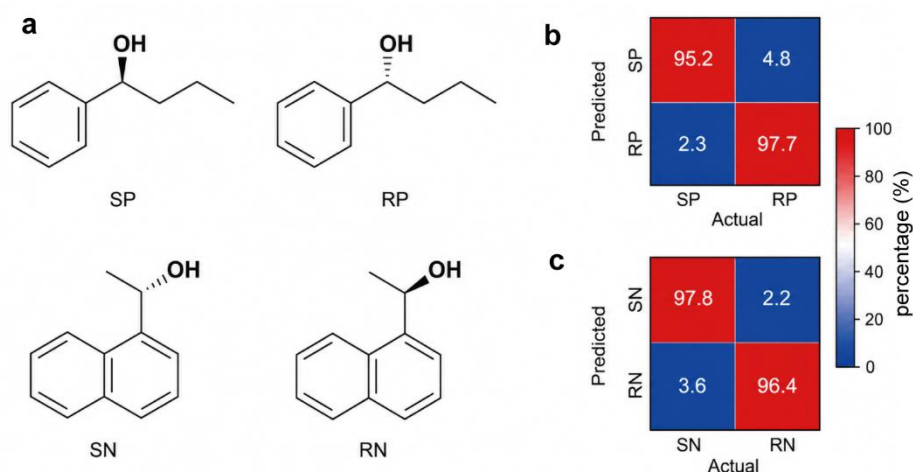
*Estimated using a total of 28 amino acids identified and quantified via LC-MS analysis. This study only reports the abundances of 11 out of these 28 species.

Supplementary Note 11. Single-molecule discrimination of chiral alcohols using nanogap electrodes

Single-molecule measurements of aqueous solutions of chiral alcohols were conducted to evaluate the discriminative capability of the proposed method for molecules beyond amino acids (Supplementary Fig. 20). The four tested alcohols (Supplementary Fig. 21) were (S)-1-phenylbutan-1-ol (SP), (R)-1-phenylbutan-1-ol (RP), (S)-1-(naphthalen-1-yl)ethan-1-ol (SN), and (R)-1-(naphthalen-1-yl)ethan-1-ol (RN). Fifteen-minute measurements yielded 288, 210, 843, and 1,178 single-molecule signal events for SN, RN, SP, and RP, respectively (Supplementary Fig. 20). For machine learning, 90% of each dataset was used for training and the remaining 10% for testing. Discrimination accuracy, evaluated using the F_1 -score, reached 96.2% for the (S)/(R) pair of 1-phenylbutan-1-ol (Supplementary Fig. 21b) and 97.0% for 1-(naphthalen-1-yl)ethan-1-ol (Supplementary Fig. 21c), respectively, comparable to those achieved for amino acids. These results demonstrate that the proposed single-molecule measurement method can be effectively applied to the identification of other chiral molecules at the single-molecule level.



Supplementary Figure 20. Current-time profiles of four chiral alcohols. Representative current-time profiles obtained from single-molecule electrical measurements of (S)-1-phenylbutan-1-ol (SP, **a**), (R)-1-phenylbutan-1-ol (RP, **b**), (S)-1-(naphthalen-1-yl)ethan-1-ol (SN, **c**), and (R)-1-(naphthalen-1-yl)ethan-1-ol (RN, **d**). Samples were dissolved in Milli-Q water at a concentration of 100 nM before measurement. Each experiment was independently repeated three times ($n = 3$ independent measurements) under identical experimental conditions, with similar results.



Supplementary Figure 21. Chiral discrimination of alcohols. **a**, Chemical structures of (S)-1-phenylbutan-1-ol (SP), (R)-1-phenylbutan-1-ol (RP), (S)-1-(naphthalen-1-yl)ethan-1-ol (SN), and (R)-1-(naphthalen-1-yl)ethan-1-ol (RN). **b**, Confusion matrix for SP and RP. The values in the confusion matrix represent the percentage (%) of current-time waveforms assigned to each class. The color scale also represents the percentage (%) of waveforms assigned to each class. **c**, Confusion matrix for SN and RN. The values in the confusion matrix represent the percentage (%) of current-time waveforms assigned to each class. The color scale also represents the percentage (%) of waveforms assigned to each class.

Supplementary Note 12. Chemical and sample materials

Amino-acid standards were commercially obtained from Tokyo Chemical Industry Co., Ltd. (TCI) or FUJIFILM Wako Pure Chemical Corporation and used as received unless otherwise stated. The chemical identity, enantiomeric form, catalogue number, stated purity, and preparation conditions for each standard are listed in Supplementary Table 1. The analysed molecular classes comprised the L- and D-enantiomers of 19 proteinogenic amino acids and glycine.

The Murchison meteorite sample was obtained from the Smithsonian National Museum of Natural History. Atacama Desert sediment samples were collected from the Diego de Almagro (DGA) and north of Antofagasta (NOA) sites in February 2022. Sample provenance, preparation, extraction procedures, and handling conditions are described in the Methods.

Supplementary Table 3. Event flow for confidence-filtered nanogap analysis of environmental extracts and the process blank.

Sample	Total picked-up events	Feature-extracted events	High-confidence target events ^a	Unknown/rejected events ^b	Target fraction ^c (%)
Murchison	41,350	39,128	19,437	19,691	49.7
Process blank	10,681	10,681	727	9,954	6.8
NOA	13,240	12,517	3,182	9,335	25.4
DGA	23,380	21,992	6,231	15,761	28.3

^aHigh-confidence target events were feature-extracted events with a maximum predicted probability ≥ 0.5 for one of the predefined 11 target species: glycine and the D- and L-enantiomers of alanine, valine, serine, aspartic acid, and glutamic acid.

^bUnknown/rejected events were feature-extracted events with a maximum predicted probability < 0.5 . These events were excluded from the confidence-filtered composition and D/L-ratio analyses.

^cTarget fraction was calculated as the number of high-confidence target events divided by the number of feature-extracted events $\times 100$.

Supplementary Table 5. Class distribution of retained high-confidence target assignments in the process blank.

Molecular species	Event count	Fraction of retained high-confidence blank events (%)
Glycine	245	33.7
D-Alanine	2	0.3
L-Alanine	41	5.6
D-Valine	4	0.6
L-Valine	52	7.2
D-Serine	0	0.0
L-Serine	3	0.4
D-Aspartic acid	0	0.0
L-Aspartic acid	373	51.3
D-Glutamic acid	0	0.0
L-Glutamic acid	7	1.0
Total	727	100.0

For the six-amino-acid group analysis, D- and L-enantiomer assignments were combined. The resulting composition was glycine, 33.7%; alanine, 5.9%; valine, 7.7%; serine, 0.4%; aspartic acid, 51.3%; and glutamic acid, 1.0%.

Supplementary References

1. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794 (Association for Computing Machinery, 2016).
2. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
3. Amashukeli, X., Pelletier, C. C., Kirby, J. P. & Grunthaner, F. J. Subcritical water extraction of amino acids from Atacama Desert soils. *J. Geophys. Res. Biogeosci.* **112**, G04S16 (2007).
4. Glavin, D. P. et al. Extraterrestrial amino acids and L-enantiomeric excesses in the CM2 carbonaceous chondrites Aguas Zarcas and Murchison. *Meteorit. Planet. Sci.* **56**, 148–173 (2021).