

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

Corresponding Author: Dr Masateru Taniguchi

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Versatile and robust chiral discrimination strategies for wide varieties of chiral compounds are of significant importance in multidisciplinary fields. Recently, many developments in chiral discrimination have been achieved through various techniques, which can be divided into two major categories based on different recognition mechanisms. First, methods rely on the inherent optical activities of chiral molecules; second, strategies depend on traditional lock-and-key molecular recognition. In the current work, the authors developed a method based on quantum electron tunneling for the discrimination of several amino acid molecules. Although the work has some merits, the method cannot realize an ideal chiral discrimination strategy. Undoubtedly, high recognition sensitivity, absolute enantiospecificity, and versatility are three prerequisites of an ideal chiral discrimination strategy, which cannot meet by the presented method. Additionally, the quantum electron tunneling strategy has been frequently used by the authors for monitoring different species (see references 23-25, 27-28). Selecting small molecules (chiral amino acids) as a new target and discriminating with a similar strategy and poor accuracy is not a significant novelty that would be the interest of Nature Communications readers. Last but not least, the developed method lacks real-world application and lacks superiority compared with existing methods. Therefore, I think that the present work cannot meet the standards of Nature Communications.

In the following specific comments are presented for the AUTHORS improving the manuscript:

1. It seems that by reporting deceiver information the authors try to convince the readers that their work has impressive benefits! For example, in lines 36-38: if the authors intended to evaluate enantiomeric excess of biologically relevant molecules, why only amino acids were selected as chiral compounds? Why carbohydrates with several chiral centers were ignored? Carbohydrates are biologically relevant molecules as important as amino acids. Another surprising point mentions the "search for life beyond Earth" (lines 37-38). From the beginning of the manuscript, I thought it is an interesting application of the chiral discrimination method. However, after reading the whole manuscript, I was disappointed, since it is routine work without even investigation of real-work applications on Earth not beyond Earth!
2. The authors claimed that "There are no robust methodologies to define the ratio of and D-form amino acids when the absolute concentration of material is extremely low." (lines 38-39). It seems that the authors ignored recently published papers in the field of chiral discrimination. There are several methods with much better performance compared with the developed method. Please see: Arabi M, Ostovan A, Wang Y, et al. Chiral molecular imprinting-based SERS detection strategy for absolute enantiomeric discrimination[J]. Nature Communications, 2022, 13(1): 5757; Liu Z, Ai J, Kumar P, et al. Enantiomeric Discrimination by Surface-Enhanced Raman Scattering—Chiral Anisotropy of Chiral Nanostructured Gold Films[J]. Angewandte Chemie, 2020, 132(35): 15338-15343; Leong S X, Koh C S L, Sim H Y F, et al. Enantiospecific molecular fingerprinting using potential-modulated surface-enhanced Raman scattering to achieve label-free chiral differentiation[J]. ACS Nano, 2021, 15(1): 1817-1825. These methods not only work for the differentiation of amino acids but also a wide variety of chiral compounds can be discriminated. As seen, the developed method lacks significant advantages over existing methods.
3. The accuracy of the developed method either for identifying individual amino acids (80 %) or blind test (50 %) is much lower than previous methods. Additionally, the developed method perfectly failed in competition with the methods which enabled absolute enantiomeric discrimination.
4. Despite the authors' claim the method even lacks sensitivity. The concentration of all tested samples was 0.10 mM (line 228), which is too high. Obviously, their method cannot use for the discrimination of amino acids at the single-molecule level. It is expected, if the lower concentrations of amino acids will be tested, this insufficient accuracy will even significantly be

reduced.

5. Can this method apply to chiral recognition as well? How about the LOD and LOQ of the method?

6. Lines 55-59,199-200: As I mentioned before, there is no experimental evidence for the claims. It seems that the authors completely ignored the fundamental of space sample analysis. There are many co-existing compounds with amino acids. How do the authors intend to analyze samples with their method? What are the details of the experimental procedure? How do they obtain the samples? Related information should be added to the manuscript.

7. I was surprised that none of the chiral discrimination strategies that have been developed until now considered by the authors! There isn't even one sentence about previous chiral discrimination strategies in the Introduction part! Accordingly, it seems that the authors are not familiar with the current status of chiral discrimination methods. So how they can overcome the recent challenges?

8. Another disappointing point is there is no information and characterization about gold nanogap electrodes. The authors simply cited their previous works without providing any information. Why 0.56 and 0.58 nm nanogaps were selected? How about batch-to-batch reproducibility of the electrode? How do they control the sharpness of the gold tips?

9. There are three main concerns about the developed method. First, lacking versatility. According to the presented mechanism, the method only works for amino acids and aqueous solutions. Current industries need chiral discrimination of a wide variety of chiral molecules not only amino acids, such as chiral drugs, alcohols, pesticides, etc. Most of the chiral drugs are not soluble in water and lack special functional groups. The ability of the developed method for the discrimination of such compounds should be studied. The second and third concerns provide in comments 14 and 16.

10. Which chemical/physical parameters affect the sensitivity and accuracy of the method? How the authors optimized the variables during nanogap device construction? Moreover, the molecular size of all target amino acids should be reported in a Table.

11. Does the orientation of amino acids between two electrodes affect accuracy? This issue should be investigated.

12. Schematic illustration in Figure 1a does not make sense. The single amino acid molecule cannot provide a signal. Also, the morphology of gold nanogap is demonstrated as agglomerated particles, which is not in accordance with the TEM image.

13. Line 111-112: as I aforementioned, the developed method cannot apply to neutral chiral.

14. Second concern: there is a vague impression about the real-world application of the current method due to a lack of systematic analysis. According to the suggested mechanism, the method is highly susceptible to matrix impurities during chiral discrimination in real samples. Many compounds, such as ions, small molecules, and even nonchiral amino acids can interference within analysis.

15. In addition, Au tips tend to etch within real-sample analysis resulting in poor accuracy.

16. Third concern: the method has negligible potential for impalement in general laboratories. HPLC-based methods provide even better discrimination efficiency, which widely uses in industries and academia. Besides, the HPLC method can be performed by an untrained operator. In contrast, the developed method needs advanced technology and highly expert operators for device construction and data analysis.

17. The chemical structure of amino acids and water molecules after electronic interaction should be investigated. Is there any change in the chemical/physical structure of amino acids and water molecules? How about side reactions and possible products?

18. The authors only considered the horizontal nanogap. However, vertical gap between the gold tips is about 200 μm . See TEM image in figure 1. So, it is impossible and uncontrollable that only single amino acid molecule stay in this gap.

19. pH of the water sample after dissolving amino acids is changed. Since there is no information about the pH of water samples, does it mean the developed method is pH independent?

20. Several unconsidered variables affect the method sensitivity/accuracy. The density charge on the surface of amino acid molecules is different. Also, by considering the molecular weight/size of the amino acid molecules the retention time of amino acid in the nanogap is completely different. The manuscript is like only an experimental report, not a research paper since it lacks such critical analysis.

21. Line 184: there is no information about the quantitative analysis of four amino acids. The analytical figure of merits should be reported for the quantitative analysis, including response range, LOD, matrix effect, and reproducibility.

22. Reproducibility of the whole method is missing as well. For example, it is not clear all measurements were performed by one device or different devices. Moreover, information related to the device stability and reusability should be added.

23. Most of the "Discussion" parts are related to the potential application, not valid applications! For each claim related experimental facts should be provided.

Reviewer #2

(Remarks to the Author)

In this manuscript, Ohshiro et al. describe how tunneling current measurements using a gold nanogap with high miller index can be used to discern all L- and D-forms of all canonical amino acids. As a possible application, the authors suggest that detecting L-favoring stoichiometry in extraterrestrial samples can serve as evidence for the existence extraterrestrial life. In experiments, the authors show that a machine learning algorithm can be trained to discern the tunneling current signals generated by all 39 amino acid enantiomers.

I am excited about the reported ability to tell enantiomers apart. However, I am left with many major concerns regarding the methodology in this paper, which I believe should be addressed by revision.

1. For the homogeneous sample analyses, were the test and training data drawn from different experiments, or were events from the same experiment divided over non-overlapping test and training data sets? This should be noted explicitly. If the latter, how can the authors be sure that the stated classification performances were achieved solely on the basis of event

properties, and not experiment-specific properties? For instance, in figure 2a and b, it is clear that even the baseline contains sufficient information to tell the two apart, so even if I would extract both test and training data from these fragments, without any events at all, I would be able to classify them perfectly. It seems that no signal normalization has been performed either to remove such artifacts.

2. In the second part, solutions of 4 amino acids are measured, to determine if the known molar ratios between the constituents can be recovered. This is a good test of the method; however, I have several issues with this part as well.

2.1. Why are these combinations of amino acids chosen? In neither of the presented cases are L- and D-variants of the same residue included, while I was under the impression that discriminating between L and D was something the method was supposed to do.

2.2. The authors are reporting identification error rates. That means that they must have a ground truth; however, if these are measurements of mixtures, how do they know the ground truth identity for each individual waveform?

2.3. The ratios of amino acids are given with an uncertainty measure, but it is not stated what this uncertainty measure is (confidence interval? prediction interval?) or how they have come by it.

2.4. Percentages are stated to be “within error” (line 164). What does this mean? By what standard or by what test?

3. In general, the analysis method is not properly explained. The type of classifier used has been left completely unmentioned; it is only referred to as “machine learning”. The methods give a hint and state that XGBoost is used – which by the way, is not part of the scikit-learn library as the authors state, but a stand-alone package. This tells us that it is a boosted tree method, but which one? I cannot check for myself, because the code has not been made available. For the data a doi link is given, but it is invalid. Nature reporting standards state that both code and data should be freely available.

4. In lines 127-139, the authors argue that a difference in F-measure when gap size is varied is evidence for their theory on the basis for enantiomer differentiation. Specifically, they state that the F-measure increased from 0.86 to 0.89 when the gap was reduced, and state that this is (statistically?) significant given the number of waveforms. I suspect that this statistical significance has been deduced incorrectly. The sd for the F-measure is not directly dependent on the number of waveforms. Rather, the model needs to be evaluated repeatedly to collect many F-measures (e.g. by bootstrapping). Then using a statistical equivalence test (for which the parameters should be noted), the significance of the found difference should be gauged.

5. The authors make an excellent point in line 174-177, i.e. that different transport features per amino acid can lead to different relative counts, but then they do not apply this realization. With the data of homogeneous samples in the first part of the manuscript, they can easily check this and if they truly find different rates, they could even correct the rates measured in heterogeneous samples.

6. Regarding novelty, the authors state that the SM identification of chiral molecules has not been previously achieved. While I tend to agree that it is not a fully solved problem, there have been several publications on this topic that I am aware of, particularly using biological nanopores [1,2]. At the very least, a mention of this work in the discussion is warranted in my opinion.

Furthermore, I note the following minor issues:

1. line 49: “... setting a new standard of sensitivity”: sensitivity has a specific meaning in machine learning (i.e. recall, true positive / total positive). I disagree that a new standard has been set in this regard, consider removing.

2. Line 98: “...contain information on only one part of the waveform”: “part of the waveform” is confusing, consider “do not contain all information contained in the waveform”.

3. Line 119: “...waveforms were machine-trained”: “machine-trained” and “machine-learned” is used multiple times throughout the manuscript, but I have never seen the expression of data being “machine-trained” before, or the use of machine learning as a verb. The classifier is trained, not the data. Consider replacing instances by “the ML classifier was trained using X data” or similar.

4. Line 120: “...F-measure...”: the F-measure should be defined by a number indicating how precision and recall are weighed. I presume they are weighed equally here, so then this measure should be called the F1-score (note the subscript 1).

5. Line 120: “(Fig. 2h)”: should be Fig 2g

6. Line 124-125: “The mean..., respectively.”: it is unclear to me what these two F-measures refer to; the latter seems to be the average reported in Table 1, but what is the first one?

7. Line 124-126: “By the binomial...more waveforms”: this seems true, but define the alpha value. Moreover, is this a relevant statement? What practical application would entail analysis of enantiomerically pure samples? And following the same reasoning (and considering how short measurement takes), I could argue that a much worse classifier could classify pure samples almost perfectly given 100 or more waveforms.

8. Line 143-146: “This value... waveforms.”: it cannot be ascertained that each of the 39 amino acids can be accurately identified if the mean F-measure is given. Consider rephrasing.

9. Line 161: “33.3%...28.2%”: these percentages are repeated in line 163 again. Consider removing here.

10. Line 182: “genertaed”: should be generated.

11. Line 228: “The sample ... mM.” This is a very high concentration in single-molecule analysis terms. Below which concentration does this method become impractical?

12. Line 235: “Signal picking”: consider rephrasing to “event detection”.

- [1] Boersma, Arnold J., and Hagan Bayley. "Continuous stochastic detection of amino acid enantiomers with a protein nanopore." *Angewandte Chemie* 124.38 (2012): 9744-9747.
- [2] Jia, Wendong, et al. "Identification of Single-Molecule Catecholamine Enantiomers Using a Programmable Nanopore." *ACS nano* 16.4 (2022): 6615-6624.

Reviewer #3

(Remarks to the Author)

Review of "Chiral Discrimination of Amino Acids at the Single-Molecule Level"
by Ohshiro et al.

First of all, let me say this is a fascinating, creative and exciting approach, and believe this work is highly worthwhile and deserving of publication. However, having said that, there are a number of fundamentally important changes that can improve this work, and which need to be made in order to make it ready for publication. There are also a number of additional suggested changes that I believe would enhance the manuscript but which are not essential for publication.

Essential Topics for Revision

In the abstract the authors make these statements: "There are no robust methodologies to define the ratio of L- and D-form of amino acids when the absolute concentration of material is extremely low".

(Note I think the word "define" should be replaced with "measure" in this context...)

The authors should remove adjectival descriptions like "extremely low". Reference 41 describes an account of measuring nanomolar level quantities of amino acids and also discriminating their chiral ratios. Many readers would think this is "extremely low"! If there is to be some prose assessment of "how low is low", in this context, the only values that are of relevance are the limits of detection described in the MacKenzie paper mentioned above, or those given in the Europa Lander Science Definition Team report. You can just reference those publications, and make concrete statements about how this work compares to the values that planetary scientists think are relevant for future missions.

Additionally, when the authors say "robust methodologies", and are implying this paper could lead to a robust methodology, the very first thing to consider in this context is the presence of salts in the sample matrix. This comes up in all the working groups on astrobiology for ocean worlds, and the Europa Lander mission study, etc. This critical aspect must be addressed in this paper. What is the effect of an unpredictable mixture of salts containing Na⁺, Mg²⁺, Ca²⁺, Cl⁻, SO₄⁻, etc. on these measurements? It seems like this would greatly affect the results. If these measurements would require all salts to be removed, then there are front end sample desalting techniques that could be used prior to analysis (and that would be OK; it just adds sample processing requirements, and that should be stated). A study of the effect of different salts on the measurements is essential here to establish relevance. But beyond that, typically for journals of the same impact factor as Nature Communications (or even lower) it is essential to provide data on a real sample (not a laboratory standard) in order to demonstrate that the chemical analysis method has relevance, and is acceptable for publication. This is the general rule for Analytical Chemistry for example, and I would expect this to be the case here as well.

Another fundamental aspect that this sensitivity topic brings up is that this paper uses solutions containing 0.1 mM amino acids for all the experiments, whereas the expected concentrations in relevant samples are much lower than that. If the relevant concentration for these measurements of "chirality of the miniscule quantities of amino acids that may be present in trace samples ..." (line 58 of Introduction) is on the order of, say, 1 nM, this means that the frequency of observing an amino acid passing through the nanowire tip junction is reduced by a factor of 100,000. How long would it take to observe even one amino acid in this sample? What is the flow rate through the junction? Or, more generally, is constant flow through the junction required? Can a sample simply be recirculated in a microfluidic loop? What if the amino acids don't pass through the junction at all? If the sample size is 1 mL (for example), how much time would be required to collect a statistically relevant data set to analyze? In order for this to be applied on a spaceflight mission, is sample concentration of very dilute samples prior to analysis required? This aspect needs to be addressed as well. For reference, by analogy, the authors could consider the published work by Lindensmith, Nadeau et al. on digital holographic microscopy. One of the drawbacks of performing traditional high-resolution microscopy on a spaceflight mission is that the sample volume that can be interrogated is extremely small, and it is not possible to search through relevant sample volumes using reasonable amounts of time (i.e. this is like searching for a needle in a haystack). For that reason, those researchers developed ways to interrogate larger fluid volumes. The situation here is different of course, but a clear analysis of the times required for these measurements, and a description of how a future system could perform measurements on a reasonable timescale, is necessary. Additionally, how would the system actually be calibrated on a mission? Would you need to take along a test set of amino acids to "retrain" the system, prior to analyzing a real sample, like shown in the paper? I think this is another fundamental thing to explain – how would these measurements actually be performed on a mission? What if the amino acids on the astrobiology target are different than the calibration set? Would the system be able to make meaningful measurements?

Another key area that requires clarification is the description of the data analysis. For non-specialists it is really not clear how you calculate the "time signal region" in Figure 2 (e,f) and what the units are. Also please explain exactly what the value in each matrix quantity represents. In Figure 2 it says the values indicate "percentage of waveforms", but percentage of waveforms that do what? A clearer explanation of these methodologies for the non-specialist (and defining "confusion matrix" for example) would greatly improve the paper.

Suggested Revisions

I suggest that the authors add the reference below to emphasize that the astrobiology community recognizes that nanopore technology would be desirable specifically for a future Orbilander mission to Enceladus, and on the timescale of that mission launch, this technology, with an addition of the measurements described here, could potentially be ready for that opportunity: The Planetary Science Journal, 2:77 (18pp), 2021 April <https://doi.org/10.3847/PSJ/abe4da>

Another thing that immediately springs to mind when reading this work: if the measurements require gold nanotips with chiral structure, and that is an enabling feature for measurements, how can you be assured that the tips won't change over the ten or more years of interplanetary cruise that the system will sit around before it is used? If a spacecraft experiences temperatures as high as 60C during transit to the outer solar system, will the gold atoms move around and change the properties of the tip? What is the temperature dependence of the tips? If this is expected then the authors could address this and say that thermal control would likely be required, or any effects could be understood during calibration in mission operations, etc.

In figure 1 I think you can remove panel (b) where you show the structures of all the different amino acids. That is common knowledge. Also the top panel (a) is very important and should be made larger.

Finally, it is quite interesting that tryptophan has seemingly unique differences in measured properties. As the most highly conjugated amino acid, it should have the largest pi electron cloud that could contribute to electron transport across the junction, and this could be part of the explanation for these observations. If there is anything you can conclude about the nature of the amino acids themselves from the measurements made, that would be very interesting indeed. Any additional information of this kind could potentially be useful for data interpretation on future missions if amino acids are present that are not included in your set of calibration compounds.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have revised the manuscript following most of my comments and the manuscript has now been largely improved. However, several issues are still required to be addressed prior to its publication:

1. Reviewer 1, Comment 5: error bars should be added to Figure 2.
2. Table 1: sulfur-containing amino acids (e.g., cysteine) have a high affinity to gold nanogap electrodes through Au-S interactions. However, in practice, it seems this interaction does not have an influence on the td (ms). Why?
3. The fundamental quantum electron tunneling strategy that is frequently used by the authors should be briefly mentioned in the introduction. Besides, the authors should clearly demonstrate, which changes in this strategy are essential for successful discrimination of chiral molecules compared with other species. In other words, the novelty of the current work in terms of technical aspects should be indicated.
4. Reviewer 1, Comment 19: PBS buffer at pH 7.4 was used. Does the type of buffer affect the discrimination performance of the method? By considering isoelectric points of amino acids, it is still not clear why pH was set at 7.4.

Reviewer #3

(Remarks to the Author)

The revised version of the manuscript is greatly improved and the authors have made many of the revisions suggested. However they still have not performed an analysis on a natural sample which I believe is essential for this publication to appear in any of the Nature journals, or any other high impact journal for that matter. (My Comment 3). It is not possible to simply "explain away" the need for verification of an analytical technique. The purpose of testing on a real sample is to test the "unknown unknowns" and it can't be replaced by laboratory simulants. To be completely transparent with the authors I will say that this has been my personal experience as a science author publishing papers on the development of new analytical techniques as well, and I think this is fair, and a legitimate test. I have never been able to publish a new technique without analyzing a real sample. This critical test shows that the system would really work in a real world scenario. Also for the record, keep in mind that this paper is intended for a journal literally named Nature. If this paper is about a new way to explore nature it should definitely examine a natural sample.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I have thoroughly reviewed the manuscript, along with the author's responses to feedback from other reviewers. I am pleased to note that the manuscript has undergone substantial improvements, and the concerns I previously had have been effectively addressed in the revised draft. Given these improvements, I wholeheartedly recommend this article for publication in its current form.

Reviewer #3

(Remarks to the Author)

Reviewer #3 (Follow-Up Comments)

I appreciate the authors' substantial efforts to address my prior concern regarding validation on natural samples. The inclusion of analyses of the Murchison meteorite and Atacama Desert samples greatly improves this manuscript. However, after reviewing the revised version carefully, several important issues remain that I believe require clarification prior to publication.

Issue 1. Forced Classification Without an "Unknown" Category

The manuscript states that:

"All detected signals in the nanogap analysis were classified into the trained categories... as no explicit 'others' class was included."

This raises a significant analytical concern. In complex environmental samples, signals originating from contaminants or untrained molecular species will necessarily be assigned to the closest trained amino acid category. This forced-classification framework can inflate apparent agreement with LC-MS and distort both compositional and D/L ratio estimates, particularly at low abundance.

I strongly recommend that the authors do these things or equivalent to clarify the situation: A) Clarify how forced classification may bias quantitative estimates. B) Report classifier confidence metrics or probability thresholds. C) Discuss how implementation of an "unknown" rejection class would affect interpretation. D) Provide blank classification composition data to show how background signals are distributed across classes.

Without this clarification, the robustness of the method in heterogeneous real-world samples remains uncertain.

Issue 2. Background Contamination and Elevated Blank Signals

The authors note elevated signal event counts in NOA and blank samples relative to LC-MS quantification and attribute this to surface or airborne contamination.

Given the high sensitivity of the nanogap platform, contamination control becomes a huge issue in terms of quantitative reliability.

Please clarify whether background subtraction was performed and whether contamination produces systematic enrichment in particular amino acid classes. For a technique proposed for planetary exploration, contamination susceptibility must be carefully quantified.

Issue 3. D/L Ratio Correlation Strength

The reported correlation between nanogap-derived and LC-MS D/L ratios yields $R^2 = 0.719$. This level of agreement is encouraging and suggests that the method is capable of recovering meaningful chiral information from natural samples. However, the correlation is still moderate, and the relatively large uncertainties observed for several amino acids (particularly those at low abundance) indicate that enantiomeric ratio determination remains sensitive to event statistics and classification uncertainty.

To better evaluate the quantitative performance of the method, it would be helpful if the authors provide additional statistical detail such as: regression slope and intercept with confidence intervals, statistical tests of deviation from unity slope, or if they give a discussion of potential systematic bias, if present. Because chiral discrimination is a central claim of the manuscript, providing this additional analysis would strengthen the quantitative interpretation of the results. But again to be clear these are very encouraging results that should be strengthened for a Nature publication.

Issue 4. Au–Thiol Interaction and Dwell Time Interpretation

In response to another reviewers comment about cysteine, the authors argue that self-assembled monolayer formation requires approximately one day, and therefore Au–thiol interactions are unlikely to influence the 30–40 ms dwell time. However, this explanation may not fully address the relevant physical timescales. Au–thiol chemisorption can occur on sub-millisecond timescales and does not require any monolayer formation. The relevant question is whether transient coordination events measurably alter dwell-time distributions. But this whole topic is quite confusing because the authors actually say in the paper that they coat the gold with silica! What is the actual situation here?

Please clarify A) whether the nanogap apex is exposed gold or coated with SiO₂, B) Whether td distributions for cysteine show long-tail behavior indicative of transient adsorption and if the gold is not coated, c) why transient Au–S coordination does not influence dwell time in this system.

This point is important because the manuscript attributes chiral discrimination to molecule–electrode electronic coupling. This entire discussion calls into question the overall mechanistic interpretation in the paper. The manuscript proposes chiral-dependent electronic coupling at atomically structured gold surfaces as the basis for discrimination. However no structural characterization of electrode chirality is presented, no theoretical modeling is provided, and the SiO₂ coating description would seem to negate all this reasoning. I recommend clarifying all of this, and tempering the mechanistic claims and

framing them as hypotheses unless supported by direct evidence.

Issue 5. Concentration Regime and Practical Sensitivity for a Real Mission

The manuscript reports an LOD of 0.03 nM based on calibration against a blank signal frequency of $2.1 \pm 0.27 \text{ min}^{-1}$.

However, after a careful analysis of their end-to-end analytical process for real samples like the ones you could encounter on a spaceflight mission, this value appears to represent an idealized instrumental detection threshold rather than a practically achievable identification limit under realistic conditions.

Several factors constrain the effective sensitivity:

1) Background-limited detection. A non-zero blank event rate establishes a statistical floor. At subnanomolar concentrations, analyte signal frequencies would approach background levels, and because the classifier lacks an “unknown” category, background events will necessarily be assigned to amino acid classes. This inflates apparent signal and limits reliable identification near the stated LOD. Under these conditions, the effective detection limit is determined by the rate at which analyte-derived events can be distinguished from background events over the stable operational lifetime of the nanogap device.

2) Inability to extend acquisition time. The authors state that prolonged measurement does not substantially increase usable events due to adsorption and nanogap instability. This prevents improving sensitivity via longer integration, fundamentally constraining the operational detection limit.

3) Matrix and validation gap. Natural samples were analyzed at micromolar concentrations which are several orders of magnitude above the claimed LOD. No data are presented demonstrating accurate identification or D/L quantification at nanomolar levels in heterogeneous matrices. Therefore, the practical detection limit for complex samples has not yet been demonstrated and cannot be inferred from the calibration-based LOD reported here.

For these reasons, the quoted LOD of 0.03 nM is not directly relevant to mission-level trace detection as currently demonstrated and should be reframed accordingly. The authors should clearly distinguish between calibration-based detection limits under controlled conditions and the practical identification limits achievable in complex environmental samples.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

.....

Chiral Discrimination of Amino Acids at the Single-molecule Level

.....

By Takahito Ohshiro, Yuki Komoto, Takayuki Takaai, Takeshi Yoshida, Takashi Washio, Daniel Duzdevich, Maria T. Zuber, Christopher E. Carr, Masateru Taniguchi

We express our sincere gratitude to the referees for their valuable and helpful comments. We have studied the comments carefully and made the necessary revisions.

We concur with the unanimous comment of the reviewers on the insufficiency of the original data and explanation of AI for this study during its initial submission. We have attempted to address the stated shortcomings, and the manuscript has been extensively revised. The following content has been added to the main text and Supporting Information:

Additional content included in revised manuscript

Addition to main text

- Correlation between signal frequency and t_d (Figure 2e)
- Dependence of signal frequency on tryptophan concentration (Figure 2f)
- Correlation between discrimination accuracy and molecular length and between discrimination accuracy and molecular weight (Figure 3c)
- Quantitative analysis of mixtures of L-His, D-His, L-Ala, L-Gln, and L-Ser (Figures 4h-4i)
- Quantitative analysis of seven mixed solutions consisting of two nucleic acids (contaminants) and five amino acids, including L and D bodies (Figures 4k and 4l)
- Chiral identification of two sets of L- and D-body alcohols (Figure 5)
- Properties of amino acids (Table 1)

Addition to Supporting Information

- Correlation of molecular weight, molecular length, isoelectric point, t_d , and signal frequency of amino acids (Figure S6)
- Machine learning flowchart (Figure S7)
- Discrimination accuracy for L-tryptophan and D-tryptophan measured using training and independent test data (hold-out method) (Figure S8)
- Current–time profiles of two sets of chiral alcohols (Figure S9)
- Discrimination accuracy for amino acids measured using reused electrodes (Figure S10)
- Time dependence of distance between nanogap electrodes (Figure S11)
- Classifier dependence of discrimination accuracy for chiral molecules (Table S1)

Reviewer #1 (Remarks to the Author):

Versatile and robust chiral discrimination strategies for wide varieties of chiral compounds are of significant importance in multidisciplinary fields. Recently, many developments in chiral discrimination have been achieved through various techniques, which can be divided into two major categories based on different recognition mechanisms. First, methods rely on the inherent optical activities of chiral molecules; second, strategies depend on traditional lock-and-key molecular recognition. In the current work, the authors developed a method based on quantum electron tunneling for the discrimination of several amino acid molecules. Although the work has some merits, the method cannot realize an ideal chiral discrimination strategy. Undoubtedly, high recognition sensitivity, absolute enantiospecificity, and versatility are three prerequisites of an ideal chiral discrimination strategy, which cannot meet by the presented method. Additionally, the quantum electron tunneling strategy has been frequently used by the authors for monitoring different species (see references 23-25, 27-28). Selecting small molecules (chiral amino acids) as a new target and discriminating with a similar strategy and poor accuracy is not a significant novelty that would be the interest of Nature Communications readers. Last but not least, the developed method lacks real-world application and lacks superiority compared with existing methods. Therefore, I think that the present work cannot meet the standards of Nature Communications.

In the following specific comments are presented for the AUTHORS improving the manuscript:

Comment 1

It seems that by reporting deceiver information the authors try to convince the readers that their work has impressive benefits! For example, in lines 36-38: if the authors intended to evaluate enantiomeric excess of biologically relevant molecules, why only amino acids were selected as chiral compounds? Why carbohydrates with several chiral centers were ignored? Carbohydrates are biologically relevant molecules as important as amino acids. Another surprising point mentions the "search for life beyond Earth" (lines 37-38). From the beginning of the manuscript, I thought it is an interesting application of the chiral discrimination method. However, after reading the whole manuscript, I was disappointed, since it is routine work without even investigation of real-work applications on Earth not beyond Earth!

Reply 1

As the reviewer points out, living organisms are composed of L-form amino acids and D-form sugars. Therefore, the chiral discrimination of sugars can also be an indicator of the existence of

living organisms. We focus on amino acids for the following reasons:

The 20 amino acids in living things have different molecular structures and compositions and therefore electronic states. This makes the 20 species easy to identify. By contrast, sugars have the same molecular composition but different steric configurations. This difference in conformation produces a slight variation in their electronic states. In other words, sugars are expected to differ in single-molecule conductance only slightly. Therefore, the signal-to-noise ratio should be improved further to identify sugars accurately. Research and development on the chirality identification of sugars is ongoing. As the reviewer expects, the single-molecule identification of sugars will lead to the development of a sugar sequencer, which does not exist yet, and revolutionize glycoscience research. To clarify the above details, we have specified amino acids and sugars in the Abstract and the Introduction and clarified that this study focuses on amino acids.

<Main text>

Between amino acids and sugars, we focused on the chiral identification of amino acids, which are commonly subjected to single-molecule measurements.

2. The authors claimed that “There are no robust methodologies to define the ratio of and D-form amino acids when the absolute concentration of material is extremely low.” (lines 38-39). It seems that the authors ignored recently published papers in the field of chiral discrimination. There are several methods with much better performance compared with the developed method. Please see: Arabi M, Ostovan A, Wang Y, et al. Chiral molecular imprinting-based SERS detection strategy for absolute enantiomeric discrimination[J]. Nature Communications, 2022, 13(1): 5757; Liu Z, Ai J, Kumar P, et al. Enantiomeric Discrimination by Surface-Enhanced Raman Scattering–Chiral Anisotropy of Chiral Nanostructured Gold Films[J]. Angewandte Chemie, 2020, 132(35): 15338-15343; Leong S X, Koh C S L, Sim H Y F, et al. Enantiospecific molecular fingerprinting using potential-modulated surface-enhanced Raman scattering to achieve label-free chiral differentiation[J]. ACS Nano, 2021, 15(1): 1817-1825. These methods not only work for the differentiation of amino acids but also a wide variety of chiral compounds can be discriminated. As seen, the developed method lacks significant advantages over existing methods.

Reply 2

We thank the reviewer for these references. We have added two papers about the use of gold nanoparticle surfaces to our references because they show results that support our findings. Furthermore, a major difference between our method and Raman spectroscopy is that our method electrically identifies chirality. To clarify this, we have revised the following. In addition,

we have changed the ambiguous phrase "extremely low" to "subnanomolar" to emphasize the difficulty of subnanomolar-scale measurement in current planetary science.

The developed electrical measurement system is portable. Optical measurement systems, such as Raman spectrometers, are not resistant to mechanical vibration. By contrast, our measurement system resists mechanical vibration without requiring an antivibration table. In fact, we have installed our single-molecule measurement system on an aircraft and performed measurements under microgravity conditions during parabolic flight (Ramírez-Colón et al. *submitted*). Distinguishing between Raman and fluorescence signals also requires high precision timing. In contrast, events detected during nanogap measurements occur with millisecond durations, making measurements less sensitive to noise and reducing measurement precision and bandwidth requirements. This simplifies the engineering design for portable systems including for applications such as life detection beyond Earth.

<Abstract>

There are no robust electrical methods for measuring the ratio of L- and D-form amino acids and sugars at subnanomolar absolute concentrations of materials.

Comment 3

The accuracy of the developed method either for identifying individual amino acids (80 %) or blind test (50 %) is much lower than previous methods. Additionally, the developed method perfectly failed in competition with the methods which enabled absolute enantiomeric discrimination.

Reply 3

The single-molecule discrimination accuracy described in this paper is for a single current–time waveform. Actual measurements require the discrimination accuracy not of a single current–time waveform but of a single sample. If the accuracy of discriminating two amino acids using a single current–time waveform is approximately 90% and more than 16 current–time waveforms are obtained from the measurement of one sample, then the discrimination accuracy for this sample is over 99.9% due to the binomial distribution. The 50% value is for cases where two amino acids cannot be distinguished, that is, random readings. To clarify this, we have added the following to the text:

<Main text>

These obtained F_1 –scores indicate the accuracy of discriminating two molecules using a single current–time waveform. In a real use case, one solution is measured to identify molecule A or B.

The 5 min measurement of a single solution yields more than several hundred current–time waveforms. If the identification accuracy with one waveform is p and n waveforms are obtained, then the identification accuracy for one solution follows the binomial distribution $\sum_{n/2}^n {}_nC_t p^t (1-p)^{n-t}$. For $p = 0.9$ and $n = 16$, the identification accuracy is 99.9%.

Comment 4

Despite the authors' claim the method even lacks sensitivity. The concentration of all tested samples was 0.10 mM (line 228), which is too high. Obviously, their method cannot use for the discrimination of amino acids at the single-molecule level. It is expected, if the lower concentrations of amino acids will be tested, this insufficient accuracy will even significantly be reduced.

Reply 4

We apologize for this error. We tested μM concentrations and have changed mM to μM .

Comment 5

Can this method apply to chiral recognition as well? How about the LOD and LOQ of the method?

Reply 5

To answer this comment and related remarks of the other reviewers, we performed additional experiments to determine the detection limit of this single-molecule measurement system and added their explanation to the manuscript.

The developed system is capable of the chiral identification of amino acids. Following the reviewer's suggestion, additional experiments were conducted to examine the detection limit of this single-molecule measurement system, and its LOD and LOQ were obtained. The results on the dependence of the signal frequency on tryptophan concentration have been added as Figure 2f, and the following has been added to the text:

<Main text>

The dependence of tryptophan detection frequency on concentration was assessed to determine the limit of detection (LOD) of the single-molecule measurement system (Figure 2f). The measurement time at each concentration was 15 min. The standard deviation of the signal frequency at each concentration was less than 0.27 min^{-1} . As the mean signal frequency of the blank was $2.1 \pm 0.27 \text{ min}^{-1}$, the LOD and limit of quantification (LOQ) were determined to be 0.03 and 0.2 nM, respectively, from the calibration curve. At high concentrations (above 10 mM), the

signal frequency no longer depended on the concentration of amino acids because the amino acids had reached saturation adsorption. A target limit of detection adopted for as a reference for the Europa Lander mission is 1 nM, and the (presumed abiotic background) abundance of amino acids in Mars meteorites are well above this concentration. Therefore, the nanogap single-molecule measurement system has sufficient detection sensitivity.

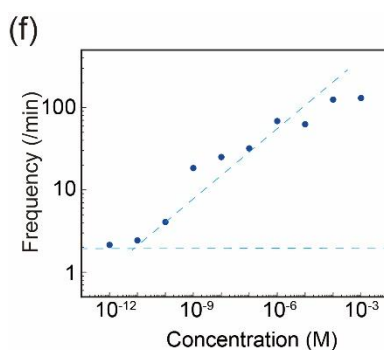


Figure 2. f, Dependence of tryptophan signal frequency on concentration. The blue dotted lines are the fitted straight lines for estimating the LOD and LOQ.

Comment 6

Lines 55-59,199-200: As I mentioned before, there is no experimental evidence for the claims. It seems that the authors completely ignored the fundamental of space sample analysis. There are many co-existing compounds with amino acids. How do the authors intend to analyze samples with their method? What are the details of the experimental procedure? How do they obtain the samples? Related information should be added to the manuscript.

Reply 6

The purpose of our work is not to describe the specific extraction method for life detection on Mars or other worlds. Our purpose is to prove the feasibility of chiral detection. Our proposed space instrument (Carr et al. *Astrobiology* 23(10) (2023), <https://www.liebertpub.com/doi/full/10.1089/ast.2022.0119>) is at a low technology readiness and part of that process involves increasingly sophisticated processing of samples. This is beyond the scope of the current manuscript, however we note that prior and ongoing space missions include drills, systems to powderize a sample, liquid chemistry, and other necessary components for sample preparation.

Comment 7

I was surprised that none of the chiral discrimination strategies that have been developed until now considered by the authors! There isn't even one sentence about previous chiral

discrimination strategies in the Introduction part! Accordingly, it seems that the authors are not familiar with the current status of chiral discrimination methods. So how they can overcome the recent challenges?

Reply 7

In accordance with this remark, a previous method of identifying optical activity has been added to the Introduction.

<Main text>

The molecules of these chiral isomers have the same chemical properties, so they generally cannot be distinguished using differences in reactivity or chemical properties. Two methods are mainly used for the chiral discrimination of samples: using polarized light to determine differences in optical properties and using molecular recognition to identify differences in conformation. In many cases, separation discrimination is performed by binding a chiral molecule to a separation medium or sensor. Recently, chiral identification has been performed via the interaction of chiral molecules with molecules modified on the walls of bionanopores.

For life detection beyond Earth, such as at Mars, Enceladus, or Europa, the measurement system must be small and light enough to be carried by a spacecraft, and resistant to vibration. Furthermore, because the number of organic molecules in extraterrestrial environments is expected to be small, the system should be able to measure a concentration of ~ 1 nM molecules, which is far lower than the amino acid abundance measured in Mars meteorites. Direct electrical detection can simplify detector design compared to optical detection, which can contribute to smaller size and potentially reduce sensitivity to vibration. To minimize the risks of false negatives and limit sample preparation complexity, it is desirable to avoid chemical or enzymatic processing steps such as labelling reactions. In other words, a measurement system that electrically detects and identifies chiral molecules at low concentrations, ideally at the single-molecule level, without using recognition molecules, is highly desirable.

Comment 8

Another disappointing point is there is no information and characterization about gold nanogap electrodes. The authors simply cited their previous works without providing any information. Why 0.56 and 0.58 nm nanogaps were selected? How about batch-to-batch reproducibility of the electrode? How do they control the sharpness of the gold tips?

Reply 8

In accordance with the reviewer's comment, the fabrication of gold nanogap electrodes is detailed in the revised Methods section as follows:

<Methods>

When the piezo element is pushed up, the thin metal wire is gradually bent and broken, forming a nanogap electrode. When the piezo element is pulled down, the nanogap electrode is bonded. Repeating the breakage and bonding processes through the up-and-down movement of the piezoelectric element gradually sharpens the electrode tip, eventually forming a single-atom contact. The formation of a gold monoatomic contact yields quantized conductance ($77.5\ \mu\text{S}$), after which the single-atom contact is broken. After this breakage, a tunneling current flows between the nanogap electrodes. The tunneling current and the distance between electrodes (d) are related by $I \propto \exp(\beta d)$. The damping constant $\beta = -4\pi\sqrt{2m\phi}/h$, where m , ϕ , and h denote the electron mass, the work function of gold electrodes and Planck constant. The distance between the nanogap electrodes is monitored through this tunneling current. If the distance deviates by $\pm 0.0075\ \text{nm}$ from the set distance, then feedback control is performed using a piezoelectric element to maintain the set distance between the electrodes. Consequently, the set distance between the nanogap electrodes is maintained at $\sigma = \pm 0.0052\ \text{nm}$ for more than 5 h (Figure S11).

<SI>

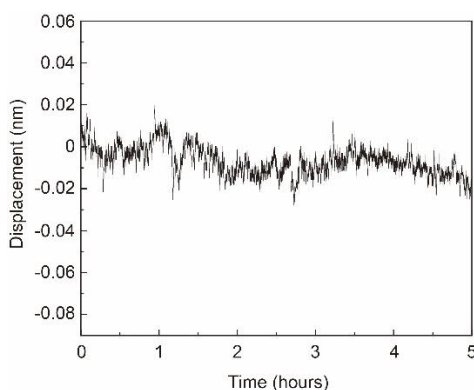


Figure S11. Time dependence of displacement between nanogap electrodes.

Comment 9

There are three main concerns about the developed method. First, lacking versatility. According to the presented mechanism, the method only works for amino acids and aqueous solutions. Current industries need chiral discrimination of a wide variety of chiral molecules not only amino acids, such as chiral drugs, alcohols, pesticides, etc. Most of the chiral drugs are not soluble in water and lack special functional groups. The ability of the developed method for the

discrimination of such compounds should be studied. The second and third concerns provide in comments 14 and 16.

Reply 9

In accordance with the reviewer's comment, chiral alcohols were identified. The obtained results are described in the Discussion section, and the confusion matrix is in Figure 5.

<Main text>

Single-molecule measurements of aqueous solutions of chiral alcohols were performed for discrimination via machine learning to evaluate the discriminative ability of the proposed method for molecules other than amino acid molecules (Figure 5a). Four alcohols were used: (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 1178 single-molecule signals for SN, RN, SP, and RP, respectively (Figure S9). Ninety percent of each molecular signal was used as training data for machine learning, and the remaining 10% was used as test data. Discrimination accuracy was evaluated using the F_1 -score. The discrimination accuracies for the (S) and (R) forms of 1-phenyl-1-butanol and 1-naphthyl-ethanol were 96.2% (Figure 5e) and 97% (Figure 5f), respectively, which were as high as that for amino acids. Therefore, the proposed single-molecule measurement method can be applied to the single-molecule identification of other chiral molecules.

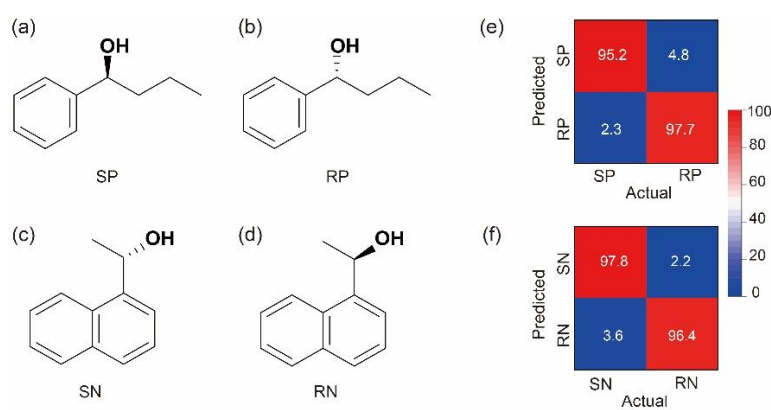


Figure 5. Chiral discrimination of alcohols. **a**, Molecular structures of (S)-1-phenylbutan-1-ol (SP), (R)-1-phenylbutan-1-ol (RP), (S)-1-(naphthalen-1-yl)ethan-1-ol (SN), and (R)-1-(naphthalen-1-yl)ethan-1-ol (RN). **b**, Confusion matrix of SP and RP. **c**, Confusion matrix of SN and RN. The matrix elements indicate the percentages of current-time waveforms out of all waveforms used in the study.

<SI>

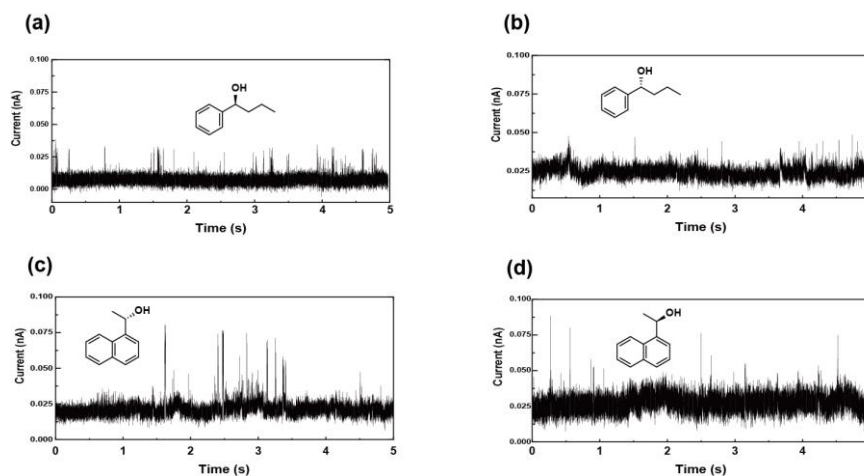


Figure S9. Current-time profiles of four alcohols. Each sample was purchased from Fujifilm Wako Pure Chemical Corporation (Tokyo, Japan) and used without further purification. Each sample was dissolved in Milli-Q water, adjusted to 100 nM, and measured. **a**, (S)-1-phenylbutan-1-ol (SP), **b**, (R)-1-phenylbutan-1-ol (RP), **c**, (S)-1-(naphthalen-1-yl)ethan-1-ol (SN), **d**, (R)-1-(naphthalen-1-yl)ethan-1-ol (RN).

Comment 10

Which chemical/physical parameters affect the sensitivity and accuracy of the method? How the authors optimized the variables during nanogap device construction? Moreover, the molecular size of all target amino acids should be reported in a Table.

Reply 10-1

The dependence of the F_1 -score on the interelectrode distance suggests that the electronic interaction between the electrode and the molecule affects the sensitivity and accuracy of the proposed method. Measurements were made at interelectrode distances of 0.52, 0.54, 0.56, 0.58, and 0.60 nm, and the discrimination accuracy decreased at 0.52 nm. For 14 of the 19 amino acids, the discrimination accuracy increased when the distance between electrodes was reduced from 0.56 nm to 0.54 nm. However, the signal frequency at the interelectrode distance of 0.54 nm was 6.5 min^{-1} , one-third that of the signal frequency at the interelectrode distance of 0.56–0.58 nm (18.4 min^{-1}). The signal frequency decreased to 4.3 min^{-1} when the distance between electrodes exceeded 0.58 nm. Therefore, an interelectrode distance of 0.56–0.58 nm was used in this experiment. To clarify this experimental condition, the following has been added to the text.

<Methods>

Measurements were performed at interelectrode distances of 0.52, 0.54, 0.56, 0.58, and 0.60 nm, where 0.52 nm resulted in low discrimination accuracy. For 14 of the 19 amino acids, the discrimination accuracy increased when the interelectrode distance reduced from 0.56–0.58 nm to 0.54 nm. However, the signal frequency at the 0.54 nm interelectrode distance was 6.5 min⁻¹, which was one-third of the signal frequency at the 0.56–0.58 nm interelectrode distance (18.4 min⁻¹). The signal frequency decreased to 4.3 min⁻¹ when the interelectrode distance exceeded 0.58 nm. Therefore, an interelectrode distance of 0.56–0.58 nm was used in this study.

Reply 10-2

In accordance with the reviewer's comment, the molecular lengths of the amino acids optimized by quantum chemical calculations have been added to Table 1. These calculations are explained in Methods. In addition, t_d and signal frequency have been added to Table 1.

<Methods>

Quantum chemical calculations. Twenty amino acids were structurally optimized using quantum chemical calculations implemented in the Gaussian 98 package. The molecular structures of all amino acids contain amino and carboxyl groups. Structural optimization was performed using density functional theory (B3LYP) and 6-31G(d, p) basis sets. The molecular length was defined as the longest interatomic distance in the molecule.

<Table>

Table 1. Properties of amino acids. IP indicates the isoelectric point. Molecular length indicates the maximum interatomic distance within a molecule.

Amino acid	3- letter symbols	Molecular length (nm)	IP	t_d (ms)		Frequency (/min)		F_1 -score
				L-form	D-form	L-form	D-form	
Alanine	Ala	67	6.00	35.8	36.5	197	108	0.814
Arginine	Arg	148	10.8	36.9	30.0	160	177	0.955
Asparagine	Asn	96	5.41	42.0	32.4	84.1	162	0.860
Aspartate	Asp	91	2.77	45.6	40.9	40.1	46.6	0.919
Cysteine	Cys	86	5.07	33.4	47.0	133	101	0.857
Glutamine	Gln	114	5.65	43.1	40.6	58.6	115	0.784
Glutamate	Glu	109	3.22	34.3	35.1	131	95.6	0.876
Histidine	His	118	7.59	44.9	30.7	111	399	0.901
Isoleucine	Ile	124	6.02	29.6	35.7	443	67.0	0.974
Leucine	Leu	124	5.98	39.2	36.1	61.1	97.2	0.890
Lysine	Lys	135	9.75	52.8	39.4	46.1	108	0.792

Methionine	Met	124	5.74	37.0	34.2	109	105	0.940
Phenylalanine	Phe	135	5.48	36.8	38.1	184	102	0.875
Proline	Pro	90	6.30	41.3	34.8	37.5	87.9	0.875
Serine	Ser	73	5.68	40.1	36.0	67.2	159	0.894
Threonine	Thr	93	6.16	42.7	44.1	60.4	68.7	0.851
Tryptophan	Trp	163	5.89	39.9	35.9	44.3	175	0.827
Tyrosine	Tyr	141	5.66	40.2	37.5	207	207	0.869
Valine	Val	105	5.96	40.9	55.4	78.6	63.4	0.883
Glycine	Gly	48	5.97	43.8		82.7		-

Comment 11

Does the orientation of amino acids between two electrodes affect accuracy? This issue should be investigated.

Reply 11

Single-molecule conductance is determined by the HOMO energy state of the molecule and the electronic interaction between the molecule and the electrode. Therefore, the steric configuration of the amino acid relative to the electrode affects the electronic interaction between the electrode and the molecule. Mainly, the electrode–molecule distance affects the electronic interaction between the electrode and the molecule. Measurements were performed by changing the interelectrode distance to 0.54 nm to investigate how much the steric configuration of the amino acid relative to the electrode affects the discrimination accuracy of the proposed method. The results are shown in Figure 2b and compared with those obtained at the interelectrode distance of 0.56 nm. The correlations between the interelectrode distance, molecular length, molecular weight, and isoelectric point at 0.54 and 0.56 nm were examined, and the results have been appended to the paper.

<Main text>

The electronic interaction between the nanogap electrode and the amino acid strongly depended on the steric configuration of the amino acid relative to the nanogap electrode, especially the distance between the electrode and the amino acid. The correlation coefficients between the discrimination accuracy and molecular length at interelectrode distances of 0.54 and 0.56 nm were 0.17 and −0.02, respectively, whereas those between the discrimination accuracy and molecular weight at interelectrode distances of 0.54 and 0.56 nm were 0.20 and 0.00, respectively (Figure 3c). The isoelectric point did not correlate with the interelectrode

distance, molecular size, or molecular weight. The discrimination accuracy at the interelectrode distance of 0.54 nm was positively correlated with the molecular length and molecular weight. Therefore, in a space with a more restricted orientation of amino acids, the larger the molecular length and molecular weight, the more susceptible the molecules are to orientation restrictions.

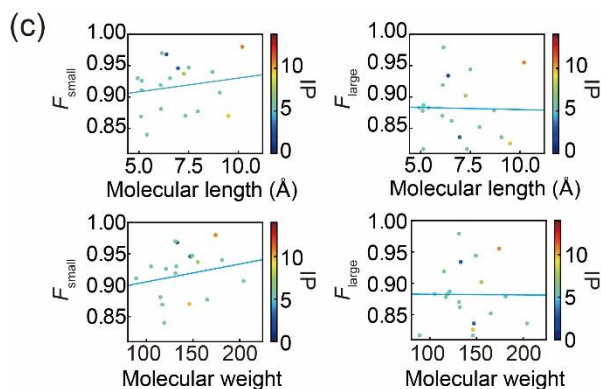


Figure 3. c. Correlations between discrimination accuracy and molecular length and between discrimination accuracy and molecular weight. F_{small} and F_{large} denote the F_1 -scores at nanogap electrode distances of 0.54 and 0.56 nm, respectively. The color scale denotes the isoelectric points (IPs) of amino acids.

Comment 12

Schematic illustration in Figure 1a does not make sense. The single amino acid molecule cannot provide a signal. Also, the morphology of gold nanogap is demonstrated as agglomerated particles, which is not in accordance with the TEM image.

Reply 12

In accordance with this comment, we have revised Figure 1a as follows:

<Main text>

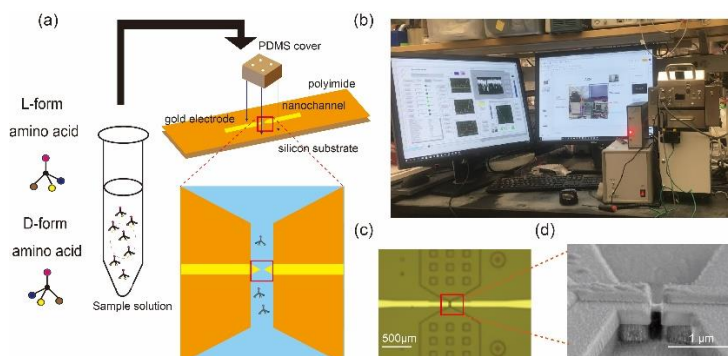


Figure 1. Schematics of chirality analysis of amino acid molecules using nanogap device. **a.** When an amino acid molecule passes between the nanogap electrodes, it acts as an electron transporter yielding a tunneling current and a measureable single-molecule electrical signal. The L-form and D-form have different electronic behavior due to the difference in the crystal orientation of the gold tip, resulting in a characteristic signal. **b,** Single-molecule measurement system. The single-molecule measurement device is on the right side of the monitor. **c,** Optical microscope image of nanogap electrode. **d,** Scanning electron microscope image of nanogap electrode.

Comment 13

Line 111-112: as I aforementioned, the developed method cannot apply to neutral chiral.

Reply 13

All measurements were performed at pH 7.4. The isoelectric points of the 39 amino acids are not 7.4, so the net charge of a single molecule is not neutral. In addition, in this experiment, the amino acid molecules were transported to the nanogap electrode mainly via diffusion. Therefore, this method can be applied to single-molecule measurements independently of the net charge of the molecule being measured.

Electrophoresis can be used to increase the frequency of amino acid detection. However, amino acids whose isoelectric points are close to the measurement pH may not be effectively transported to the nanogap electrode via electrophoresis. To solve this issue, we developed a method that uses electroosmotic flow in combination with electrophoresis. The flow channel material is SiO₂, whose solid surface is negatively charged. Therefore, cations accumulate at the solid–liquid interface of SiO₂. The electric field for electrophoresis causes cations to flow, and this flow becomes electroosmotic flow, which is then the driving force that transports amino acids regardless of their charge. We demonstrated DNA transport through electrophoresis and electroosmotic flow in a fused nanogap electrode–channel structure (Micromachines, 11, (2020) 982). However, electrophoresis requires high signal-to-noise ratios and fast measurement. Technologies incorporating electrophoresis are currently under research and development. The following statements have been added to the text to clarify that the nanostructures can use electroosmotic flow and that amino acids are transported to the nanogaps via diffusion in this study. In addition, the fabrication of fused structures of SiO₂ nanofluidic channels and thin metal wires has been added to Methods.

<Main text>

The gold nanogap electrodes were coated with SiO₂ to reduce electrical noise.

In this experiment, amino acids were transported to the nanogap electrode via diffusion.

<Methods>

A 5 μM -thick polyimide film was formed on a 0.3 mm-thick silicon substrate. An electron beam resist was spin-coated on the polyimide film, and a fine metal line pattern was drawn through electrolithography. Then, 5 nm Cr and 200 nm Au were deposited via RF sputtering. The metal wires were fabricated through the lift-off process. SiO_2 with a thickness of 300 nm was deposited via RF sputtering. After the spin-coating of the electron beam resist on the SiO_2 surface, the channel pattern was drawn via electron beam lithography. SiO_2 channels were formed through reactive ion etching using CHF_3 as the reaction gas. Finally, the polyimide directly under the metal wires was etched via reactive ion etching using oxygen as the reaction gas. The notch in the polyimide just below the metal wires makes it easy to break the metal wire.

Comment 14

Second concern: there is a vague impression about the real-world application of the current method due to a lack of systematic analysis. According to the suggested mechanism, the method is highly susceptible to matrix impurities during chiral discrimination in real samples. Many compounds, such as ions, small molecules, and even nonchiral amino acids can interfere within analysis.

Reply 14

As mentioned above, a use case is to measure a filtrate filtered through a 30 nm membrane filter using this single-molecule measurement system. The gap between the nanogap electrodes is less than 1 nm, so substances larger than 1 nm cannot be measured. In other words, the nanogap electrodes act as a physical filter. However, as the reviewer points out, contaminants smaller than 1 nm may exist. For example, assuming the existence of organisms on the Martian surface, a sample containing organisms has amino acids, nucleic acids, and sugars. Therefore, nucleic acids and sugars become contaminants. Here, we regarded the DNA monomers guanosine (dGMP) and thymine (dTMP) as contaminants. We then performed additional experiments to identify the molecules of dGMP, dTMP, glycine (Gly), and the L and D bodies of alanine (Ala) and histidine (His). The results showed an F_1 -score (identification accuracy) of 0.761. Furthermore, our quantitative analysis of seven mixed solutions based on the single-molecule measurements showed that the mixing-ratio error for each molecule was less than 33%. Thus, amino acid molecules can be identified even in the presence of contaminants. These additional results have been added to the text as follows:

<Main text>

We considered the following use case of the proposed method. We assumed the presence of organisms in sand obtained from the Martian surface. Here, we regarded the nucleic acid monomers guanosine (dGMP) and thymine (dTMP) as contaminants. We performed the single-molecule identification of seven kinds of nucleic acid monomers: dGMP, dTMP, glycine (Gly), and the L and D bodies of alanine (Ala) and histidine (His). The F_1 -score for the seven species was 0.761. The mixing ratios were quantitatively analyzed by measuring mixed solutions with a D-His:L-His:D-Ala:L-Ala:Gly:dGMP:dTMP molar ratio of 15:15:23:23:23:1:1. The single-molecule measurement result was D-His:L-His:D-Ala:L-Ala:Gly:dGMP:dTMP = 16:10:28:18:25:1:1. The error in the mixing ratio of each molecule was less than 33%. An actual sample is expected to contain more contaminants. We will prepare for the on-site detection and identification of amino acids by creating a classifier that learns as much as possible from measured data of expected contaminants.

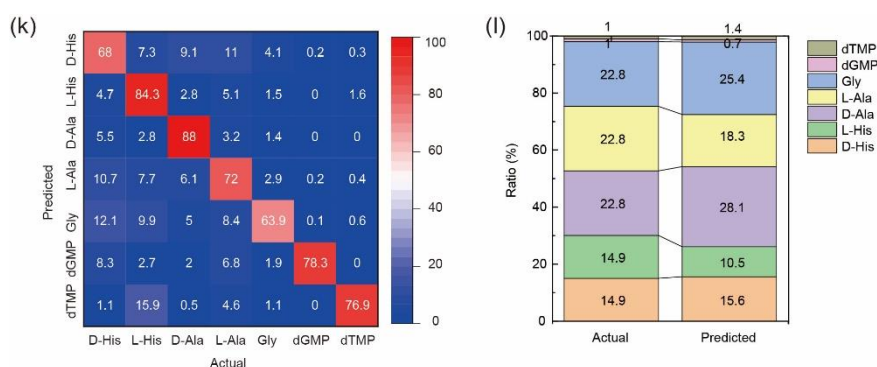


Figure 4. **k.** Confusion matrix of D-His, L-His, D-Ala, L-Ala, Gly, dGMP, and dTMP. **l.** Quantitative analysis of mixed solutions of D-His, L-His, D-Ala, L-Ala, Gly, dGMP, and dTMP.

Comment 15

In addition, Au tips tend to etch within real-sample analysis resulting in poor accuracy.

Reply 15

We thank the reviewer for this important experimental point. We did not obtain any evidence that the tips of the Au nanogap electrodes were etched by the material. However, without feedback control, the distance between the Au nanogap electrodes will indeed fluctuate. As mentioned above, the developed single-molecule measurement system monitors the tunneling current between the nanogap electrodes and uses a piezoelectric element to provide feedback control to maintain the set distance between the nanogap electrodes. This feedback control reduces the variation in single-molecule identification accuracy due to the displacement of the

interelectrode distance. This explanation has been added to Methods.

Comment 16

Third concern: the method has negligible potential for impalement in general laboratories. HPLC-based methods provide even better discrimination efficiency, which widely uses in industries and academia. Besides, the HPLC method can be performed by an untrained operator. In contrast, the developed method needs advanced technology and highly expert operators for device construction and data analysis.

Reply 16

We thank the reviewer for pointing out these issues related to real use cases. Our single-molecule measurement system was initially a special instrument that only experimental experts could measure. However, producing a large amount of homogeneous data using a special device was extremely difficult. Therefore, we developed a process of fabricating nanogap devices with quality control. We also made the single-molecule measurement system automatic to enable anyone to perform measurements easily. Now, a part-time researcher can simply place a sample in the measurement system and press the start button. Moreover, the size of the measurement device was reduced for portability; it is currently the size of an A4 sheet. To clarify the operation of the single-molecule measurement system, we have added a photograph of the measurement device in Figure 1 and specified in the Methods section that the measurement is automatic.

<Methods>

For measurement, the operator places the solution to be measured into the nanogap chip, as shown in Figure 1. The nanogap chip is placed in the measurement device, the start button is pressed, and measurement begins. First, a nanogap with an interelectrode distance preset by the operator automatically forms. Then, sample measurement starts automatically. During measurement, the distance between the nanogap electrodes is automatically feedback controlled and held at the set value. Measurement is automatically terminated when the measurement time entered by the operator is reached.

Comment 17

The chemical structure of amino acids and water molecules after electronic interaction should be investigated. Is there any change in the chemical/physical structure of amino acids and water molecules? How about side reactions and possible products?

Reply 17

We thank the reviewer for this interesting suggestion. Current technology does not enable the direct examination of changes (if any) in the amino acids and water molecules that interact electronically with the electrode. Considering the bias voltage of 0.1 V and the redox potential of amino acids and water, the occurrence of a redox reaction is unlikely. Moreover, because the molecular vibrations of amino acid and water molecules are not excited at an energy of 0.1 V, no change in electronic states due to electron injection into the molecules is considered to have occurred. We conducted experiments on single-molecule junctions in which a single molecule was chemically coupled to an electrode through an Au–S bond. However, no change in single-molecule conductance was observed at a bias of 0.1 V (Nat. Commun. 1 (2010) 138, Chem. Sci. 1 (2010), 247).

Comment 18

The authors only considered the horizontal nanogap. However, vertical gap between the gold tips is about 200 μm . See TEM image in figure 1. So, it is impossible and uncontrollable that only single amino acid molecule stay in this gap.

Reply 18

As the reviewer points out, a single molecule can enter the nanogap electrode from both vertical and horizontal directions. Molecules with an intramolecular charge, such as nucleic and amino acids, have dipole moments. Due to the strong interaction between the strong electric field and the dipole moment between the nanogap electrodes, a molecule has a stable conformation with respect to the electrodes. Simulations of nucleic acids support this prediction (M. Zwolak and M. Di Ventra, Rev. Mod. Phys., 80 (2008) 141).

The number of molecules electronically bonded between nanogap electrodes has long been debated in the study of single-molecule conductance measurements. In experiments on molecular junctions via Au–S bonds, when N molecules are junctioned between electrodes, a conductance of N times the single-molecule conductance is observed (Chem. Rev. 116 (2016) 4318). In other words, conductance histograms show mode frequencies corresponding to one, two, and three molecules. Many simulations support this experimental result. In the present experiment, the conductance histogram of the amino acids reveals a single mode (Chem. Rev. 116 (2016) 4318). Therefore, we can conclude that a single-molecule junction is formed.

Comment 19

pH of the water sample after dissolving amino acids is changed. Since there is no information about the pH of water samples, does it mean the developed method is pH independent?

Reply 19

In this study, 1× PBS buffer at pH 7.4 was used to maintain a constant pH. To clarify this, the following has been added to the manuscript:

<Main text>

Because amino acids have different isoelectric points, measurements must be made at the same pH. A 1× PBS solution at pH 7.4 was used to maintain a constant pH of the sample solution.

Comment 20

Several unconsidered variables affect the method sensitivity/accuracy. The density charge on the surface of amino acid molecules is different. Also, by considering the molecular weight/size of the amino acid molecules the retention time of amino acid in the nanogap is completely different. The manuscript is like only an experimental report, not a research paper since it lacks such critical analysis.

Reply 20

As pointed out by the reviewer, amino acids have different isoelectric points and therefore different net charges. In addition, they have different molecular lengths and weights. The time a molecule stays at the nanogap electrode, t_d , may depend on these three physical quantities and may correlate with the signal frequency. We have added a summary of the isoelectric point, molecular length, molecular weight, t_d , and signal frequency to Table 1. Furthermore, we examined the correlation between these five parameters and found a negative correlation between t_d and signal frequency. We explain these correlations in the revised text and SI.

<Main text>

Amino acids with different isoelectric points have different net charges, molecular lengths, and molecular weights. These physical quantities may affect their flow dynamics. In other words, these physical quantities may affect t_d and signal frequency. The correlations between the isoelectric point, molecular length, molecular weight, t_d , and signal frequency were therefore investigated (Figures 2e and S6), and a clear negative correlation was observed between t_d and signal frequency. Let us assume a model in which amino acids adsorbed on the nanogap electrode surface diffuse across the nanogap electrode surface to reach the sensing site. In this model, molecules with larger t_d , that is, those that are less likely to move on the electrode surface, are less likely to be observed. The experimental results are consistent with this model. This result suggests the involvement of the nanogap surface and indicates that chiral sites influence the nanogap electrode surface.

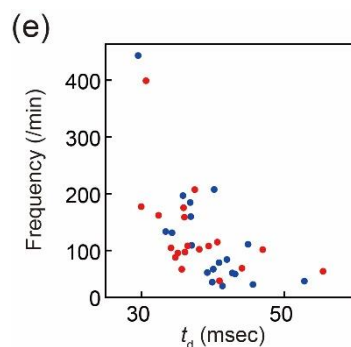


Figure 2. **e**, Correlation between signal frequency and t_d . The blue and red plots correspond to L- and D-form amino acids, respectively.

<SI>

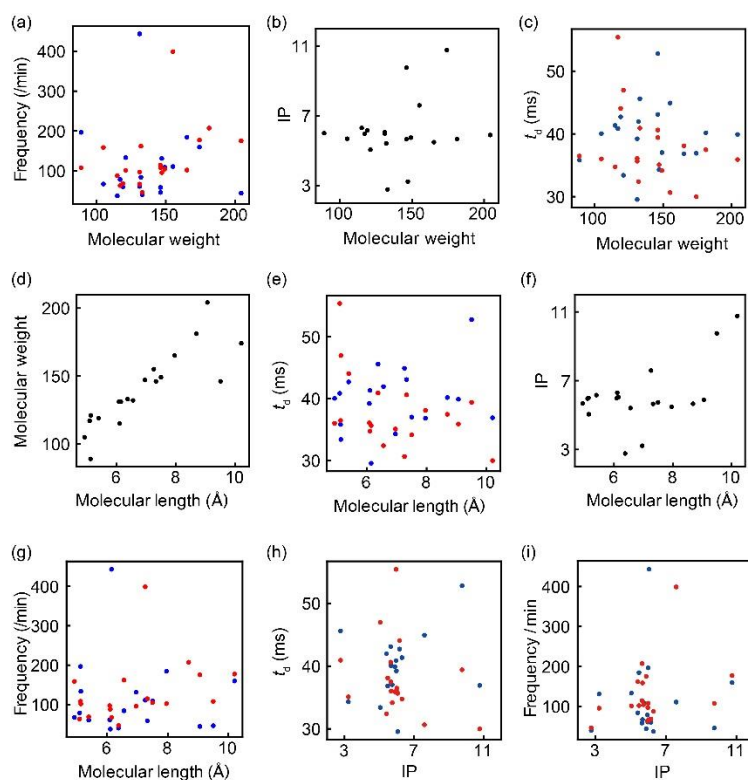


Figure S6. Correlation between molecular properties, signal frequency, and t_d . Correlation between molecular weight and **a**, signal frequency, **b**, IP, and **c**, t_d . Correlation between molecular length and **d**, molecular weight, **e**, t_d , **f**, IP, and **g**, signal frequency. Correlation between IP and **h**, t_d and **i**, signal frequency. The correlation coefficients were 0.19, 0.22, -0.13, 0.86, -0.12, 0.54, 0.15, -0.013, and 0.18, respectively. The blue and red plots indicate L- and D-form amino acids, respectively.

Comment 21

Line 184: there is no information about the quantitative analysis of four amino acids. The analytical figure of merits should be reported for the quantitative analysis, including response range, LOD, matrix effect, and reproducibility.

Reply 21

In accordance with the reviewer's comment, additional experiments were performed to clarify the response range, LOD, LOQ, and reproducibility. As mentioned in Reply 5, we have added a figure showing the concentration dependence of signal frequency to the main text.

Comment 22

Reproducibility of the whole method is missing as well. For example, it is not clear all measurements were performed by one device or different devices. Moreover, information related to the device stability and reusability should be added.

Reply 22

As described in the experimental section, each measurement was conducted using three separate nanogap devices.

As mentioned above, the distance between the nanogap electrodes is automatically feedback controlled, which can maintain the set distance between nanogap electrodes for more than 5 h. After use, the nanogap electrodes can be reused after being washed thoroughly using ultrapure water. However, cleaning may not completely flush the material from the nanogap electrode and the channel. In particular, training data should be pure amino acid data. Therefore, all nanogap electrodes in this experiment were disposed of. The following statement has been added to Methods to clarify the above. In addition, as stated in Reply 8, we have added an explanation of the time dependence of the interelectrode distance to Supporting Information. In addition, examples of measurement data obtained using reused nanogap electrodes have been added to Supporting Information.

<Method>

The nanogap electrodes can be reused after being washed using ultrapure water (Figure S10). Pure amino acid data should be utilized as training and test data to reveal the principle accuracy of the single-molecule identification of amino acids. However, a substance may not be completely washed out of the nanogap device. Therefore, in this study, the nanogap electrodes were disposed of. Nonetheless, results confirmed that the reuse of the nanogap device did not

significantly change the discrimination accuracy (Figure S10).

<Supporting Information>

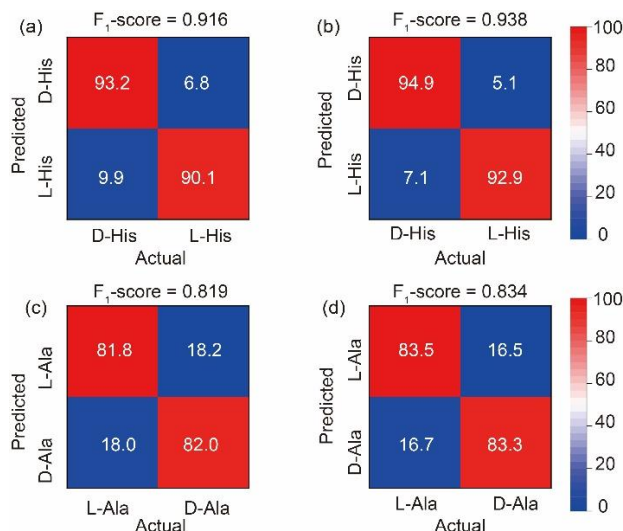


Figure S10. Identification of chiral molecules using reused nanogap electrodes. The discrimination of the L and D forms of alanine (Ala) and histidine (His) was performed using the same devices. After the first measurement, the nanogap electrodes were washed using ultrapure water. A second measurement was performed using the washed nanogap electrodes. Machine learning was implemented using the data sets obtained from the first and second measurements. The first and second discrimination accuracies for His were **a**, 0.916 and **b**, 0.950, respectively, and those for Ala were **c**, 0.819 and **d**, 0.834, respectively. In conclusion, the nanogap electrodes are reusable.

Comment 23

Most of the “Discussion” parts are related to the potential application, not valid applications! For each claim related experimental facts should be provided.

Reply 23

Following this comment, we have added the assumed pretreatment method to Discussion. In addition, we have added the results of the chiral identification of amino acids when the foreign substances are assumed to be nucleic acids, assuming the existence of organisms on the Martian surface.

Reviewer #2 (Remarks to the Author):

In this manuscript, Ohshiro et al. describe how tunneling current measurements using a gold nanogap with high miller index can be used to discern all L- and D-forms of all canonical amino acids. As a possible application, the authors suggest that detecting L-favoring stoichiometry in extraterrestrial samples can serve as evidence for the existence extraterrestrial life. In experiments, the authors show that a machine learning algorithm can be trained to discern the tunneling current signals generated by all 39 amino acid enantiomers.

I am excited about the reported ability to tell enantiomers apart. However, I am left with many major concerns regarding the methodology in this paper, which I believe should be addressed by revision.

Comment 1

For the homogeneous sample analyses, were the test and training data drawn from different experiments, or were events from the same experiment divided over non-overlapping test and training data sets? This should be noted explicitly. If the latter, how can the authors be sure that the stated classification performances were achieved solely on the basis of event properties, and not experiment-specific properties? For instance, in figure 2a and b, it is clear that even the baseline contains sufficient information to tell the two apart, so even if I would extract both test and training data from these fragments, without any events at all, I would be able to classify them perfectly. It seems that no signal normalization has been performed either to remove such artifacts.

Reply 1

In this experiment, the current–time waveform of one amino acid was obtained from three independent measurements; that is, three nanogap chips were used. A single data set was formed using the obtained data and then divided randomly and equally into 10 data sets. Nine of the 10 data sets were utilized as training data and one was used as test data for machine learning. This cross-validation method allowed us to reduce the influence of measurement-dependent artifacts. However, as the reviewer points out, in real-world use cases, training is performed on independent data sets to create a high-accuracy classifier. With this classifier implemented, the single-molecule measurement system is then transported to space or a planet, where measurements are made. This means that training and test data are obtained from independent measurements. This process, called the hold-out method, is a real use case. Therefore, we conducted an experiment to obtain training data for the L and D forms of

tryptophan and an additional experiment to obtain test data. The obtained F_1 -score (discrimination accuracy) was 0.859, which was the same as that of the cross-validation method. The above results have been added to the main text.

<Main text>

The measured data for the D and L forms of tryptophan constituted one data set each. For reducing the training bias, the data in the two data sets were undersampled, and the number of waveforms in each data set was standardized to the same number. The classifiers were trained using nine of 10 random partitions of each data set. The one remaining data segments were used to test the classifiers. The F_1 -score was calculated using the cross-validation method, which was executed 10 times.

Classifier performance is evaluated using the cross-validation and hold-out methods. The cross-validation method is commonly used to reduce training bias. However, in real-world use cases, training and test data are completely separate. For example, a classifier is trained using amino acid data measured on Earth, and a single-molecule measurement system implementing the trained classifier is transported to the Martian surface. Samples obtained from the Mars regolith are measured to determine which amino acids are present. Training and independent test data for the D and L forms of tryptophan measured using the different nanogap chips were prepared to determine the discrimination accuracy using the hold-out method, where the training and test data are completely separate. The numbers of D- and L-form waveforms in the training data were 7511 and 6158, respectively, whereas those in the test data were 5970 and 4889, respectively. Undersampling was performed in both the training and test processes. The hold-out method gave F_1 -scores of 0.816 and 0.859 for the training and test processes, respectively (Figure S8). Thus, the cross-validation and hold-out methods gave similar discrimination accuracies. Here, the discrimination accuracy was evaluated using the cross-validation method.

<SI>

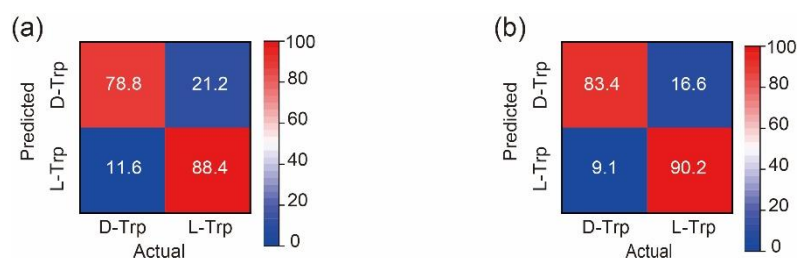


Figure S8. Discrimination accuracy for chiral molecules calculated using hold-out method. The

training and test data sets were obtained using different chips. **a**, Confusion matrix obtained in training process. **b**, Confusion matrix obtained in test process. The matrix elements indicate the percentages of obtained current–time waveforms out of all waveforms.

Comment 2

In the second part, solutions of 4 amino acids are measured, to determine if the known molar ratios between the constituents can be recovered. This is a good test of the method; however, I have several issues with this part as well.

Comment 2.1.

Why are these combinations of amino acids chosen? In neither of the presented cases are L- and D-variants of the same residue included, while I was under the impression that discriminating between L and D was something the method was supposed to do.

Reply 2.1.

In accordance with the reviewer's comments, five species (L-His, D-His, L-Ala, L-Gln, and L-Ser) were identified, and the results have been added to the text.

<Main text>

Next, the identification and quantitative analysis of five amino acids, including L and D bodies, were performed (Fig. 4h and Table 2). The identification accuracy for the five amino acids was 0.736. Mixed solutions containing L-His, D-His, L-Ala, L-Gln, and L-Ser were used. The total amino acid concentration of D-His, L-His, L-Ala, L-Gln, and L-Ser was 100 nM. Aqueous solutions with D-His:L-His:L-Ala:L-Gln:L-Ser mixing ratios of 1:1:1:1:1 and 3:1:1.5:1.5:1.5 were prepared. The 1238 and 786 current–time waveforms obtained from each measurement were classified into D-His, L-His, L-Ala, L-Gln, and L-Ser via machine learning. The numbers of classified waveforms were determined for mixing ratios of 1:1.0:0.8:0.8:1.0 (Fig. 4i) and 3:1.0:0.6:0.9:1.4 (Fig. 4j) for 1:1:1:1:1 and 3:1: 1.5:1.5:1.5. The mixing-ratio error rate of the solution with equal molarities was within 25%. However, in the 3:1:1.5:1.5:1.5 solution, L-Ala gave a maximum mixing-ratio error rate of 56%. This large error rate could be attributed to the small number of obtained current–time waveforms.

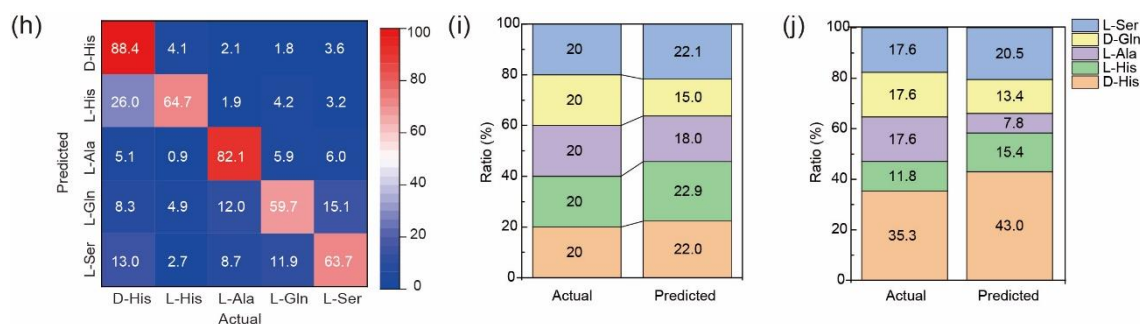


Figure 4. **h.** Confusion matrix of L-His, D-His, L-Ala, L-Gln, and L-Ser. **i.** Quantitative analysis of 1:1:1:1 D-His:L-His:L-Ala:L-Gln:L-Ser mixture. **j.** Quantitative analysis of 3:1:1.5:1.5 D-His:L-His:L-Ala:L-Gln:L-Ser mixture.

Comment 2.2.

The authors are reporting identification error rates. That means that they must have a ground truth; however, if these are measurements of mixtures, how do they know the ground truth identity for each individual waveform?

Reply 2.2.

Even with today's state-of-the-art technology, no direct measurement method can identify a single molecule that has passed through a nanogap electrode. Our ultimate goal is to develop a direct measurement method for single-molecule identification. We have developed single-molecule conductance measurement and inelastic tunneling spectroscopy (IETS), which enables single-molecule vibrational spectroscopy at low temperatures without using light (Nat. Commun. 1 (2010) 138, Chem. Sci. 1 (2010), 247). Single-molecule conductance and IETS can identify single molecules but require a low environmental temperature of 4.2 K. An embedded system for IETS is currently under research and development.

The ground truth in the mixed-solution experiment was the mixing ratio of each amino acid. Therefore, the estimated mixing ratios of the mixed solutions were expected to be comparable to the ground truth.

Comment 2.3.

The ratios of amino acids are given with an uncertainty measure, but it is not stated what this uncertainty measure is (confidence interval? prediction interval?) or how they have come by it.

2.4. Percentages are stated to be "within error" (line 164). What does this mean? By what standard or by what test?

Reply 2.3.

In accordance with this comment, the definition of the error rate has been added to the text.

<Main text>

The error rates, defined as $(\text{true value} - \text{estimated value}) / \text{true value} \times 100$, were 11.3%, 5.6%, 7.6%, and 9.6%, respectively.

Comment 3

In general, the analysis method is not properly explained. The type of classifier used has been left completely unmentioned; it is only referred to as “machine learning”. The methods give a hint and state that XGBoost is used – which by the way, is not part of the scikit-learn library as the authors state, but a stand-alone package. This tells us that it is a boosted tree method, but which one? I cannot check for myself, because the code has not been made available. For the data a doi link is given, but it is invalid. Nature reporting standards state that both code and data should be freely available.

Reply 3-1

We thank the reviewer for this comment. The adopted classifiers were a random forest, gradient boosting, an SMV, a neural network, AdaBoost, logistic regression, and a decision tree. Naïve Bayes, k-nearest neighbor, and XGBoost were used for machine learning. XGBoost was implemented in scikit-learn. Among these classifiers, XGBoost gave the highest F_1 -score. Following the reviewer’s comments, we have added the details of the machine learning method to Methods and the F_1 -scores obtained using the flowchart and the classifiers to SI.

<Methods>

The current profile was measured using an MCBJ setup. In the analysis process, pulse signals were extracted from I - t profiles (Figure S7 and Table S1). A signal is considered to have occurred if the maximum current is more than six times the standard deviation of the baseline of the signal. The signals were converted into 15-dimensional feature vectors. The features were the maximum current, the current from the baseline, duration, and 12-dimensional current factors. The 12-dimensional current factors were the average current of 12 regions partitioned along the time axis. Then, random undersampling was performed to avoid training on imbalanced data. The number of feature vectors of each species was reduced to the minimum signal number. The data set was partitioned into 10 sections and then trained and classified via 10-fold cross-validation. One section was utilized as test data for classification, and the others were for training. The XGBoost classifier was trained on the training data, and the test data were classified using the learned classifier. The accuracies were determined using the average of 10 rounds of

classification.

<SI>

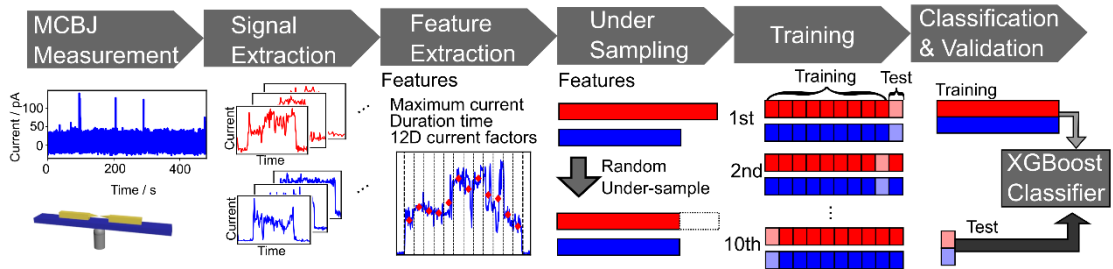


Figure S7. Scheme of machine learning analysis for chiral and amino acid discrimination.

Table S1. Classifier dependence of L-His and D-His discrimination accuracy. L-His and D-His were measured using 0.54 nm nanogap electrodes. The numbers of data obtained were 12540 and 23877, respectively. Machine learning was performed according to the flow shown in Figure S7.

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
Naïve Bayes	0.734
K-nearest neighbor	0.683

Reply3-2

We thank the reviewer for mentioning the invalid link. The access rights for Zenodo were set incorrectly. The measurement data can now be accessed on the said repository.

Comment 4

In lines 127-139, the authors argue that a difference in F-measure when gap size is varied is evidence for their theory on the basis for enantiomer differentiation. Specifically, they state that the F-measure increased from 0.86 to 0.89 when the gap was reduced, and state that this is

(statistically?) significant given the number of waveforms. I suspect that this statistical significance has been deduced incorrectly. The sd for the F-measure is not directly dependent on the number of waveforms. Rather, the model needs to be evaluated repeatedly to collect many F-measures (e.g. by bootstrapping). Then using a statistical equivalence test (for which the parameters should be noted), the significance of the found difference should be gauged.

Reply 4

As stated by the reviewer, the standard deviation of the F_1 -score is not related to the number of waveforms. Our argument is that the obtained individual F_1 -scores are sufficiently accurate because they are based on a sufficient number of current-time waveforms. Furthermore, averaging the individual F_1 -scores is incorrect. Considering the above, we have made the following revision:

<Main text>

For 14 of the 19 amino acids, the accuracy of discriminating L and D forms using the 0.54 nm nanogap was higher than that using the 0.56 nm nanogap. The Thr and Pro discrimination accuracies were higher when the 0.56 nm nanogap was used. The discrimination accuracies for Val, Met, and Ile were similar for both nanogaps. The sensitivity of the discrimination accuracy to the interelectrode distance indicates that the electronic interaction between electrodes and molecules contributes significantly to chiral discrimination.

Comment 5

The authors make an excellent point in line 174-177, i.e. that different transport features per amino acid can lead to different relative counts, but then they do not apply this realization. With the data of homogeneous samples in the first part of the manuscript, they can easily check this and if they truly find different rates, they could even correct the rates measured in heterogeneous samples.

Reply 5

As pointed out by the reviewer, amino acids have different isoelectric points and therefore different net charges. In addition, they have different molecular lengths and weights. The time a molecule stays at the nanogap electrode, t_d , may depend on these three physical quantities and may correlate with the signal frequency. We have added a summary of the isoelectric point, molecular length, molecular weight, t_d , and signal frequency to Table 1. Furthermore, we examined the correlation between these five parameters and found a negative correlation between t_d and signal frequency. We explain these correlations in the revised text and SI.

<Main text>

Amino acids with different isoelectric points have different net charges, molecular lengths, and molecular weights. These physical quantities may affect their flow dynamics. In other words, these physical quantities may affect t_d and signal frequency. The correlations between the isoelectric point, molecular length, molecular weight, t_d , and signal frequency were therefore investigated (Figures 2e and S6), and a clear negative correlation was observed between t_d and signal frequency. Let us assume a model in which amino acids adsorbed on the nanogap electrode surface diffuse across the nanogap electrode surface to reach the sensing site. In this model, molecules with larger t_d , that is, those that are less likely to move on the electrode surface, are less likely to be observed. The experimental results are consistent with this model. This result suggests the involvement of the nanogap surface and indicates that chiral sites influence the nanogap electrode surface.

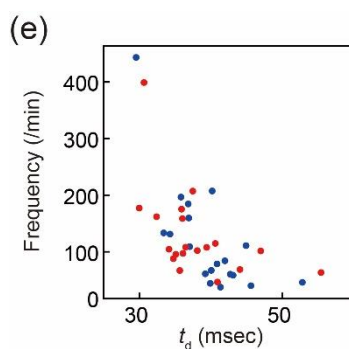


Figure 2. **e**, Correlation between signal frequency and t_d . The blue and red plots correspond to L- and D-form amino acids, respectively.

<SI>

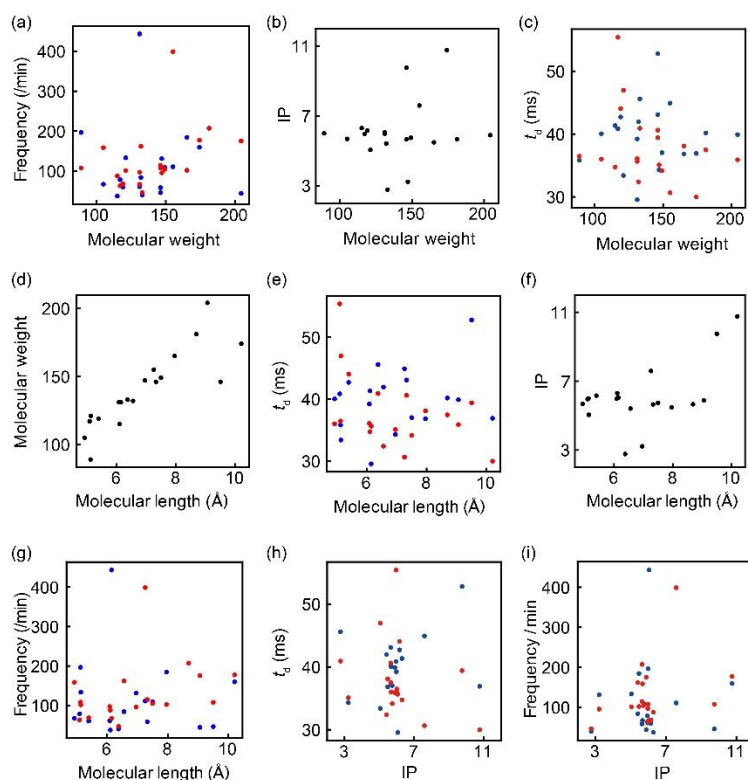


Figure S6. Correlation between molecular properties, signal frequency, and t_d . Correlation between molecular weight and **a**, signal frequency, **b**, IP, and **c**, t_d . Correlation between molecular length and **d**, molecular weight, **e**, t_d , **f**, IP, and **g**, signal frequency. Correlation between IP and **h**, t_d and **i**, signal frequency. The correlation coefficients were 0.19, 0.22, -0.13, 0.86, -0.12, 0.54, 0.15, -0.013, and 0.18, respectively. The blue and red plots indicate L- and D-form amino acids, respectively.

Comment 6

Regarding novelty, the authors state that the SM identification of chiral molecules has not been previously achieved. While I tend to agree that it is not a fully solved problem, there have been several publications on this topic that I am aware of, particularly using biological nanopores [1,2]. At the very least, a mention of this work in the discussion is warranted in my opinion.

Reply 6

We thank the reviewer for sharing these important references with us. They have been added to the Introduction. We also explain in the revised Introduction that single-molecule chiral identification is performed through the interaction between a molecule modified with a bionanopore wall surface and a chiral molecule. In addition, to clarify the novelty of this chiral identification method, we have added the identification method for other chiral molecules.

<Main text>

The molecules of these chiral isomers have the same chemical properties, so they generally cannot be distinguished using differences in reactivity or chemical properties. Two methods are mainly used for the chiral discrimination of samples: using polarized light to determine differences in optical properties and using molecular recognition to identify differences in conformation, including chirally-selective reactions. In many cases, separation discrimination is performed by binding a chiral molecule to a separation medium or sensor. Recently, chiral identification has been performed via surface-enhanced Raman scattering (SERS) and the interaction of chiral molecules with molecules modified on the walls of bionanopores.

Chiral discrimination has been proposed as a biosignature in the search for life beyond Earth. While a threshold of >20% enantiomeric excess in multiple amino acids has been proposed as indicative of life, analyses of meteoritic amino acids and sugars have revealed enantiomeric excess up to 60% for L-aspartic acid and L-glutamic acid⁵. Thus, it is desirable to characterize chiral discrimination for different amino acids and to couple chiral measurements with other evidence of biogenicity.

For life detection beyond Earth, such as at Mars, Enceladus, or Europa, the measurement system must be small and light enough to be carried by a spacecraft, and resistant to vibration. Furthermore, because the number of organic molecules in extraterrestrial environments is expected to be small, the system should be able to measure a concentration of ~1 nM molecules, which is far lower than the amino acid abundance measured in Mars meteorites. Direct electrical detection can simplify detector design compared to optical detection, which can contribute to smaller size and potentially reduce sensitivity to vibration. To minimize the risks of false negatives and limit sample preparation complexity, it is desirable to avoid chemical or enzymatic processing steps such as labelling reactions. In other words, a measurement system that electrically detects and identifies chiral molecules at low concentrations, ideally at the single-molecule level, without using recognition molecules, is highly desirable.

Furthermore, I note the following minor issues:

Comment 1. line 49: "..., setting a new standard of sensitivity": sensitivity has a specific meaning in machine learning (i.e. recall, true positive / total positive). I disagree that a new standard has been set in this regard, consider removing.

Reply 7

In accordance with the reviewer's comment, we have removed "setting a new standard of

sensitivity."

2. Line 98: "...contain information on only one part of the waveform": "part of the waveform" is confusing, consider "do not contain all information contained in the waveform".

Reply 8

We thank the reviewer for this suggestion. We have revised it to "do not contain all information contained in the waveform."

Comment 3. Line 119: "...waveforms were machine-trained" : "machine-trained" and "machine-learned" is used multiple times throughout the manuscript, but I have never seen the expression of data being "machine-trained" before, or the use of machine learning as a verb. The classifier is trained, not the data. Consider replacing instances by "the ML classifier was trained using X data" or similar.

Reply 9

We thank the reviewer for this recommendation. In accordance with the reviewer's recommendation, the following revisions have been made:

<Main text>

The machine learning classifiers were trained using the extracted waveforms.

The machine learning classifiers were trained using 3800 extracted current–time waveforms.

Comment 4. Line 120: "...F-measure...": the F-measure should be defined by a number indicating how precision and recall are weighed. I presume they are weighed equally here, so then this measure should be called the F₁-score (note the subscript 1).

Reply 10

As noted by the reviewer, the measure is indeed the F₁-score. This has been corrected throughout the paper.

Comment 5. Line 120: "(Fig. 2h)": should be Fig 2g

Reply 11

We are thankful for this correction. It has been applied.

Comment 6. Line 124-125: “The mean...,respectively.”: it is unclear to me what these two F-measures refer to; the latter seems to be the average reported in Table 1, but what is the first one?

Reply 12

We thank the reviewer for pointing this out. The first value is the F_1 -score, and we have revised it as follows:

<Main text>

The F_1 -scores for all chiral identification were more than 0.757 (Figure 3a and Table 1).

Comment 7. Line 124-126: “By the binomial...more waveforms”: this seems true, but define the alpha value. Moreover, is this a relevant statement? What practical application would entail analysis of enantiomerically pure samples? And following the same reasoning (and considering how short measurement takes), I could argue that a much worse classifier could classify pure samples almost perfectly given 100 or more waveforms.

Reply 13

We thank the reviewer for this remark. Indeed, with a long measurement time, a large number of waveforms can be obtained and a sample identification accuracy close to 100% can be achieved, even with a less accurate classifier. A real use case is determining which amino acid a sample is. The following has been added to the text to clarify the above:

<Mani text>

These obtained F_1 -scores indicate the accuracy of discriminating two molecules using a single current-time waveform. In a real use case, one solution is measured to identify molecule A or B. The 5 min measurement of a single solution yields more than several hundred current-time waveforms. If the identification accuracy with one waveform is p and n waveforms are obtained, then the identification accuracy for one solution follows the binomial distribution $\sum_{t=0}^n \binom{n}{t} p^t (1-p)^{n-t}$. For $p = 0.9$ and $n = 16$, the identification accuracy is 99.9%.

Comment 8. Line 143-146: “This value... waveforms.”: it is cannot be ascertained that each of the 39 amino acids can be accurately identified if the mean F-measure is given. Consider rephrasing.

Reply 14

As stated by the reviewer, claiming that they can be identified accurately is an oversimplification.

Therefore, the following revision has been made:

<Main text>

Therefore, each of the 39 tested amino acids can be identified using individual current–time waveforms.

Comment 9. Line 161: “33.3%...28.2%”: these percentages are repeated in line 163 again. Consider removing here.

Reply 15

We thank the reviewer for pointing this out. We have deleted the repeated figures.

Comment 10. Line 182: “genertaed”: should be generated.

Reply 16

We thank the reviewer for this correction. The typographical error has been changed.

Comment 11. Line 228: “The sample ... mM.” This is a very high concentration in single-molecule analysis terms. Below which concentration does this method become impractical?

Reply 17

We apologize for this error. We have changed mM to μM . Following the reviewer’s suggestion, additional experiments were conducted to examine the detection limit of this single-molecule measurement system, and the LOD and LOQ were obtained. The results on the dependence of the signal frequency on tryptophan concentration have been added as Figure 2f, and the following has been added to the text:

<Main text>

The dependence of tryptophan detection frequency on concentration was assessed to determine the limit of detection (LOD) of the single-molecule measurement system (Figure 2f). The measurement time at each concentration was 15 min. The standard deviation of the signal frequency at each concentration was less than 0.27 min^{-1} . As the mean signal frequency of the blank was $2.1 \pm 0.27 \text{ min}^{-1}$, the LOD and limit of quantification (LOQ) were determined to be 0.03 and 0.2 nM, respectively, from the calibration curve. At high concentrations (above 10 mM), the signal frequency no longer depended on the concentration of amino acids because the amino acids had reached saturation adsorption. A target limit of detection adopted for as a reference

for the Europa Lander mission is 1 nM, and the (presumed abiotic background) abundance of amino acids in Mars meteorites are well above this concentration. Therefore, the nanogap single-molecule measurement system has sufficient detection sensitivity.

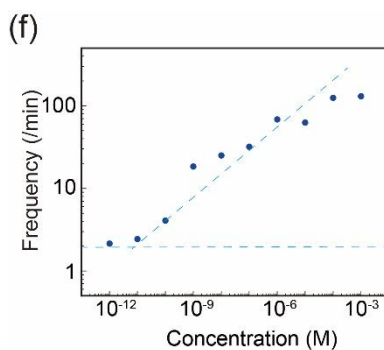


Figure 2. **f**, Dependence of tryptophan signal frequency on concentration. The blue dotted lines are the fitted straight lines for estimating the LOD and LOQ.

Comment 12. Line 235: “Signal picking”: consider rephrasing to “event detection”.

Reply 18

We thank the reviewer for pointing this out. "Signal picking" has been changed to "event detection."

Reviewer #3 (Remarks to the Author):

*Review of “Chiral Discrimination of Amino Acids at the Single-molecule Level”
by Ohshiro et al.*

First of all, let me say this is a fascinating, creative and exciting approach, and believe this work is highly worthwhile and deserving of publication. However, having said that, there are a number of fundamentally important changes that can improve this work, and which need to be made in order to make it ready for publication. There are also a number of additional suggested changes that I believe would enhance the manuscript but which are not essential for publication.

Essential Topics for Revision

Comment 1

In the abstract the authors make these statements: “There are no robust methodologies to define the ratio of L- and D-form of amino acids when the absolute concentration of material is extremely low”.

(Note I think the word “define” should be replaced with “measure” in this context...)

Reply 1

The word "define" has been changed to "measure" in accordance with the reviewer's comment.

Comment 2

The authors should remove adjectival descriptions like “extremely low”. Reference 41 describes an account of measuring nanomolar level quantities of amino acids and also discriminating their chiral ratios. Many readers would think this is “extremely low”! If there is to be some prose assessment of "how low is low", in this context, the only values that are of relevance are the limits of detection described in the MacKenzie paper mentioned above, or those given in the Europa Lander Science Definition Team report. You can just reference those publications, and make concrete statements about how this work compares to the values that planetary scientists think are relevant for future missions.

Reply 2-1

We thank the reviewer for mentioning this error, and we apologize. We have changed mM to μ M. Following the reviewer's suggestion, additional experiments were conducted to examine the detection limit of this single-molecule measurement system, and the LOD and LOQ were

obtained. The results on the dependence of the signal frequency on tryptophan concentration have been added as Figure 2f, and the following has been added to the text:

<Main text>

The dependence of tryptophan detection frequency on concentration was assessed to determine the limit of detection (LOD) of the single-molecule measurement system (Figure 2f). The measurement time at each concentration was 15 min. The standard deviation of the signal frequency at each concentration was less than 0.27 min^{-1} . As the mean signal frequency of the blank was $2.1 \pm 0.27 \text{ min}^{-1}$, the LOD and limit of quantification (LOQ) were determined to be 0.03 and 0.2 nM, respectively, from the calibration curve. At high concentrations (above 10 mM), the signal frequency no longer depended on the concentration of amino acids because the amino acids had reached saturation adsorption. A target limit of detection adopted for as a reference for the Europa Lander mission is 1 nM, and the (presumed abiotic background) abundance of amino acids in Mars meteorites are well above this concentration. Therefore, the nanogap single-molecule measurement system has sufficient detection sensitivity.

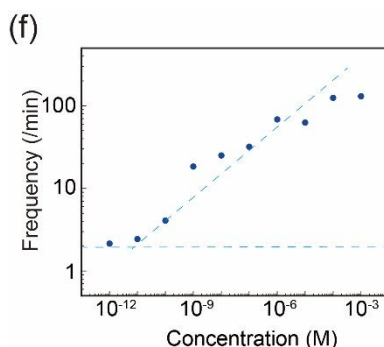


Figure 2. f, Dependence of tryptophan signal frequency on concentration. The blue dotted lines are the fitted straight lines for estimating the LOD and LOQ.

Reply 2-2

The LOD of this single-molecule measurement system is 0.03 nM. However, as the reviewer states, "extremely low" is an ambiguous term. Based on the current status of planetary science, we have made the following revision:

<Main text>

There are no robust electrical methods for defining the ratio of L- and D-form amino acids and sugars at subnanomolar absolute concentrations of materials.

Comment 3

Additionally, when the authors say “robust methodologies”, and are implying this paper could lead to a robust methodology, the very first thing to consider in this context is the presence of salts in the sample matrix. This comes up in all the working groups on astrobiology for ocean worlds, and the Europa Lander mission study, etc. This critical aspect must be addressed in this paper. What is the effect of an unpredictable mixture of salts containing Na⁺, Mg²⁺, Ca²⁺, Cl⁻, SO₄⁻, etc. on these measurements? It seems like this would greatly affect the results. If these measurements would require all salts to be removed, then there are front end sample desalting techniques that could be used prior to analysis (and that would be OK; it just adds sample processing requirements, and that should be stated). A study of the effect of different salts on the measurements is essential here to establish relevance. But beyond that, typically for journals of the same impact factor as Nature Communications (or even lower) it is essential to provide data on a real sample (not a laboratory standard) in order to demonstrate that the chemical analysis method has relevance, and is acceptable for publication. This is the general rule for Analytical Chemistry for example, and I would expect this to be the case here as well.

Reply3

Amino acids have different isoelectric points. Therefore, the electronic state of an amino acid depends on the pH of the aqueous solution. Single-molecule conductance reflects the electronic state of the amino acid. Therefore, the solution pH must be kept constant to identify the amino acid. In this experiment, the pH is maintained at 7.4 using a PBS buffer solution. Since the first manuscript did not specify that PBS was used, information about PBS and pH has been added to the text.

PBS buffer contains Na⁺, K⁺, Cl⁻, and PO₄²⁻. As the reviewer explains, the interaction of divalent cations, such as Mg²⁺ and Ca²⁺, with amino acids can be important. In particular, divalent cations and amino acids can form complexes. These complexes will be identified in our next study. However, as the reviewer points out, the measurement of real samples is necessary. As mentioned above, a use case is to measure a filtrate filtered through a 30 nm membrane filter using this single-molecule measurement system. The gap between the nanogap electrodes is less than 1 nm, so substances larger than 1 nm cannot be measured. In other words, the nanogap electrodes act as a physical filter. However, as the reviewer points out, contaminants smaller than 1 nm may exist. For example, assuming the existence of organisms on the Martian surface, a sample containing organisms has amino acids, nucleic acids, and sugars. Therefore, nucleic acids and sugars become contaminants. Here, we regarded the DNA monomers guanosine (dGMP) and thymine (dTMP) as contaminants. We then performed additional experiments to identify the molecules of dGMP, dTMP, glycine (Gly), and the L and D bodies of alanine (Ala) and

histidine (His). The results showed an F_1 -score (identification accuracy) of 0.761. Furthermore, our quantitative analysis of seven mixed solutions based on the single-molecule measurements showed that the mixing-ratio error for each molecule was less than 33%. Thus, amino acid molecules can be identified even in the presence of contaminants. These additional results have been added to the text as follows:

<Main text>

We considered the following use case of the proposed method. We assumed the presence of organisms in sand obtained from the Martian surface. Here, we regarded the nucleic acid monomers guanosine (dGMP) and thymine (dTMP) as contaminants. We performed the single-molecule identification of seven kinds of nucleic acid monomers: dGMP, dTMP, glycine (Gly), and the L and D bodies of alanine (Ala) and histidine (His). The F_1 -score for the seven species was 0.761. The mixing ratios were quantitatively analyzed by measuring mixed solutions with a D-His:L-His:D-Ala:L-Ala:Gly:dGMP:dTMP molar ratio of 15:15:23:23:23:23:1:1. The single-molecule measurement result was D-His:L-His:D-Ala:L-Ala:Gly:dGMP:dTMP = 16:10:28:18:25:1:1. The error in the mixing ratio of each molecule was less than 33%. An actual sample is expected to contain more contaminants. We will prepare for the on-site detection and identification of amino acids by creating a classifier that learns as much as possible from measured data of expected contaminants.

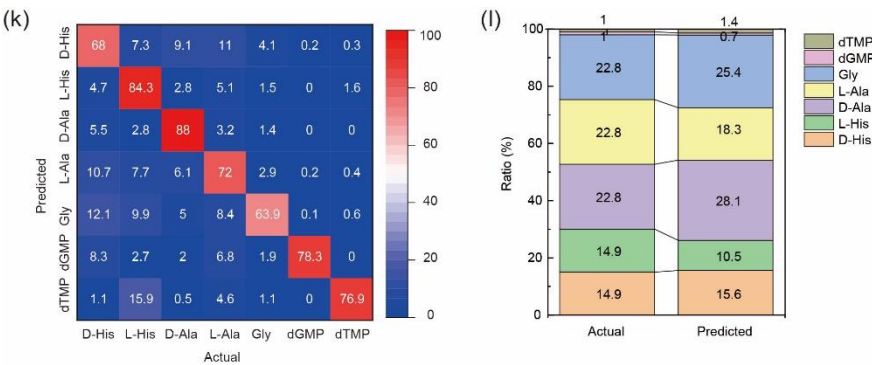


Figure 4. **k.** Confusion matrix of D-His, L-His, D-Ala, L-Ala, Gly, dGMP, and dTMP. **l.** Quantitative analysis of mixed solutions of D-His, L-His, D-Ala, L-Ala, Gly, dGMP, and dTMP.

Comment 4

Another fundamental aspect that this sensitivity topic brings up is that this paper uses solutions containing 0.1 mM amino acids for all the experiments, whereas the expected concentrations in relevant samples are much lower than that. If the relevant concentration for these measurements of “chirality of the miniscule quantities of amino acids that may be present in

trace samples ...” (line 58 of Introduction) is on the order of, say, 1nM, this means that the frequency of observing an amino acid passing through the nanowire tip junction is reduced by a factor of 100,000. How long would it take to observe even one amino acid in this sample? What is the flow rate through the junction? Or, more generally, is constant flow through the junction required? Can a sample simply be recirculated in a microfluidic loop? What if the amino acids don't pass through the junction at all? If the sample size is 1 mL (for example), how much time would be required to collect a statistically relevant data set to analyze? In order for this to be applied on a spaceflight mission, is sample concentration of very dilute samples prior to analysis required? This aspect needs to be addressed as well. For reference, by analogy, the authors could consider the published work by Lindensmith, Nadeau et al. on digital holographic microscopy. One of the drawbacks of performing traditional high-resolution microscopy on a spaceflight mission is that the sample volume that can be interrogated is extremely small, and it is not possible to search through relevant sample volumes using reasonable amounts of time (i.e. this is like searching for a needle in a haystack). For that reason, those researchers developed ways to interrogate larger fluid volumes. The situation here is different of course, but a clear analysis of the times required for these measurements, and a description of how a future system could perform measurements on a reasonable timescale, is necessary. Additionally, how would the system actually be calibrated on a mission? Would you need to take along a test set of amino acids to “retrain” the system, prior to analyzing a real sample, like shown in the paper? I think this is another fundamental thing to explain – how would these measurements actually be performed on a mission? What if the amino acids on the astrobiology target are different than the calibration set? Would the system be able to make meaningful measurements?

Reply 4-1

First, we apologize for this error. The reviewer’s major doubt stems from the mention of 0.1 mM, which is an error. We have changed mM to μ M. Additional experiments on the concentration dependence of the detection frequency were performed to clarify the LOD and LOQ of the system, and the results have been added to the text.

Reply4-2

We have added actual use cases to the Discussion section.

For example, for measurements on the Martian surface, the machine learning classifier is trained and tested using current–time waveforms of amino acids, nucleic acids, and sugars as possible residual molecules of organisms before being loaded on the spacecraft. Thus, no retraining is performed on the Martian surface. We conducted additional experiments involving nucleic acids. As mentioned above, the obtained results have been added to the text.

As pointed out by the reviewer, the challenge is measurement in the presence of unexpected molecules. This can be addressed by obtaining as much data as possible on the residues of the expected organisms.

Comment 5

Another key area that the requires clarification is the description of the data analysis. For non-specialists it is really not clear how you calculate the "time signal region" in Figure 2 (e,f) and what the units are. Also please explain exactly what the value in each matrix quantity represents. In Figure 2 it says the values indicate "percentage of waveforms", but percentage of waveforms that do what? A clearer explanation of these methodologies for the non-specialist (and defining "confusion matrix" for example) would greatly improve the paper.

Reply 5

Following the reviewer's comments, we have added the "signal time region" calculation method to Methods.

The numbers in the confusion matrix indicate the percentages of waveforms classified into each matrix element out of all waveforms. For the reader's understanding, an explanation of the confusion matrix has been added to the text. The procedure of molecular identification using machine learning has been added to Figure caption, Methods and SI.

<Figure caption>

The definition of the "signal time region" is given below. The signal time is the time between the start and end points of a signal. A signal is considered to have occurred if it is more than six times the standard deviation of the baseline of the signal. The starting point of the signal is the time the signal exceeds one standard deviation of the baseline for the first time in retrospect. The end point of the signal is the first time the signal falls below one standard deviation of the baseline. The signal time is divided into 12 segments, and the average current for each segment, or the conductance relative to the peak signal current, is shown as the value on the horizontal axis in Figures 2e–2f.

<Methods>

The current profile was measured using an MCBJ setup. In the analysis process, pulse signals were extracted from I – t profiles (Figure S7 and Table S1). A signal is considered to have occurred if the maximum current is more than six times the standard deviation of the baseline of the signal. The signals were converted into 15-dimensional feature vectors. The features were the maximum current, the current from the baseline, duration, and 12-dimensional current factors.

The 12-dimensional current factors were the average current of 12 regions partitioned along the time axis. Then, random undersampling was performed to avoid training on imbalanced data. The number of feature vectors of each species was reduced to the minimum signal number. The data set was partitioned into 10 sections and then trained and classified via 10-fold cross-validation. One section was utilized as test data for classification, and the others were for training. The XGBoost classifier was trained on the training data, and the test data were classified using the learned classifier. The accuracies were determined using the average of 10 rounds of classification.

<SI>

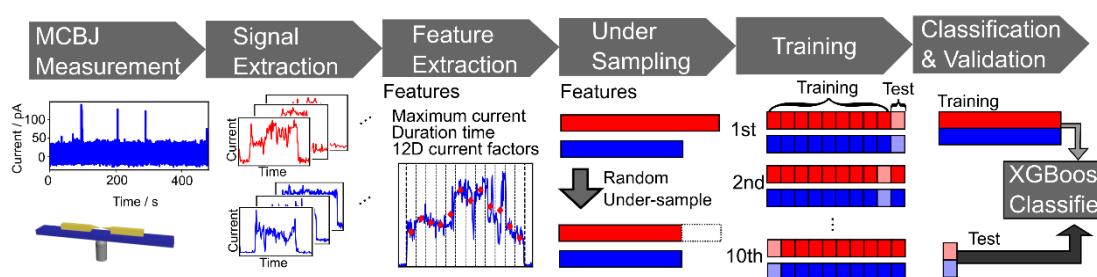


Figure S7. Scheme of machine learning analysis for chiral and amino acid discrimination.

Suggested Revisions

Comment 6

I suggest that the authors add the reference below to emphasize that the astrobiology community recognizes that nanopore technology would be desirable specifically for a future Orbilander mission to Enceladus, and on the timescale of that mission launch, this technology, with an addition of the measurements described here, could potentially be ready for that opportunity:

The Planetary Science Journal, 2:77 (18pp), 2021 April <https://doi.org/10.3847/PSJ/abe4da>

Reply 6

The following article has been added to the references in accordance with the reviewer's suggestion:

Mackenzie, S. M., et al. The Enceladus Orbilander Mission Concept: Balancing Return and Resources in the Search for Life. Planet. Sci. J. 2, 77 (2021).

Comment 7

Another thing that immediately springs to mind when reading this work: if the measurements require gold nanotips with chiral structure, and that is an enabling feature for measurements, how can you be assured that the tips won't change over the ten or more years of interplanetary cruise that the system will sit around before it is used? If a spacecraft experiences temperatures as high as 60C during transit to the outer solar system, will the gold atoms move around and change the properties of the tip? What is the temperature dependence of the tips? If this is expected then the authors could address this and say that thermal control would likely be required, or any effects could be understood during calibration in mission operations, etc.

Reply 7

As the reviewer points out, atoms near the tip of a gold nanogap electrode can migrate across the surface. However, nanogap electrodes are fabricated via mechanical breakage at the spot where they are measured. Therefore, a temperature of 60°C should not affect them. The materials used for the nanogap electrode chip are polyimide (which is heat resistant up to over 300°C), silicon, and silicon oxide, so the nanogap electrode chip is sufficiently heat resistant. To clarify this, the following has been added to the Methods:

<Methods>

Nanogap electrodes are fabricated in situ where they are measured. The polyimide, silicon, and silicon oxide that constitute the nanogap electrode chip are heat resistant up to over 300°C. Thus, the nanogap electrode chip can be transported to its destination without requiring cooling, even inside a spacecraft, where temperatures can rise to 60°C.

Comment 8

In figure 1 I think you can remove panel (b) where you show the structures of all the different amino acids. That is common knowledge. Also the top panel (a) is very important and should be made larger.

Reply 8

Figure 1b has been removed, and Figure 1a has been revised and enlarged.

Comment 9

Finally, it is quite interesting that tryptophan has seemingly unique differences in measured properties. As the most highly conjugated amino acid, it should have the largest pi electron cloud that could contribute to electron transport across the junction, and this could be part of the

explanation for these observations. If there is anything you can conclude about the nature of the amino acids themselves from the measurements made, that would be very interesting indeed. Any additional information of this kind could potentially be useful for data interpretation on future missions if amino acids are present that are not included in your set of calibration compounds.

Reply 9

We thank the reviewer for this suggestion. Indeed, tryptophan gives the highest single-molecule conductance due to its large π -electron system. We have added the following to the text:

<Main text>

Of the 20 amino acids, tryptophan has the largest π -electron system. Because a π -electron system can easily transport electrons, tryptophan has the highest single-molecule conductance.

.....
Chiral Discrimination of Amino Acids at the Single-molecule Level
.....

By Takahito Ohshiro, Yuki Komoto, Takayuki Takaai, Takeshi Yoshida, Takashi Washio, Daniel Duzdevich, Maria T. Zuber, Christopher E. Carr, Masateru Taniguchi

Following the reviewers' comments, we have revised the manuscript accordingly. Specifically, we conducted a collaborative analysis with NASA on samples from the Atacama Desert, which serves as a terrestrial analog of Mars, as well as on a meteorite from the Smithsonian Institution's collection. The results of this analysis have been incorporated into the main text and the Supporting Information.

Reviewer #1 (Remarks to the Author):

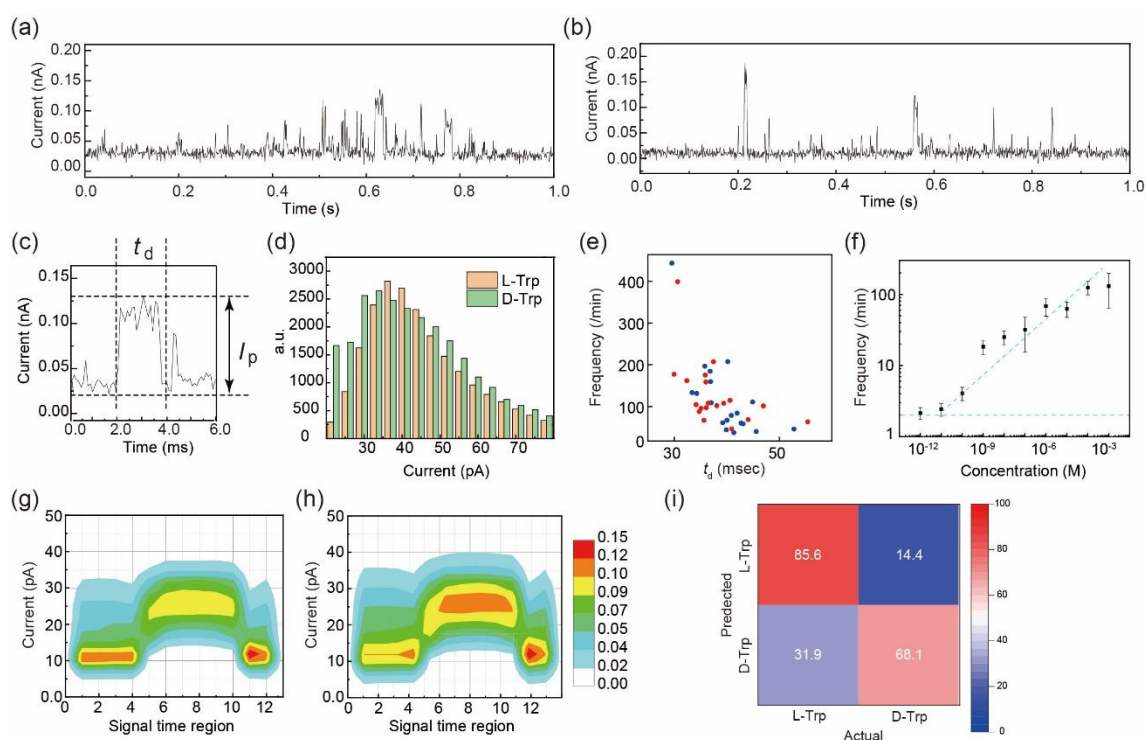
The authors have revised the manuscript following most of my comments and the manuscript has now been largely improved. However, several issues are still required to be addressed prior to its publication:

Thank you for your positive feedback. We have revised the manuscript in accordance with your comments.

1. Reviewer 1, Comment 5: error bars should be added to Figure 2.

<Reply1>

In response to your specific suggestion, we have added error bars to Figure 2f.



f, Dependence of tryptophan signal frequency on concentration. The blue dotted lines are the fitted straight lines for estimating the LOD and LOQ. The error bars represent the standard deviation divided by the square root of the number of measurements ($n = 8$).

2. *Table 1: sulfur-containing amino acids (e.g., cysteine) have a high affinity to gold nanogap electrodes through Au-S interactions. However, in practice, it seems this interaction does not have an influence on the t_d (ms). Why?*

<Reply2>

We agree with the reviewer's observation regarding the strong affinity between gold (Au) and thiol groups. In our experiments, a single cysteine molecule translocates through the nanogap within a very short duration (30–40 ms). Given that the formation of a self-assembled monolayer (SAM) via Au–thiol interactions typically requires about a day, it is unlikely that a sufficiently strong interaction would form within this brief transit time. The transit time, t_d , is a critical parameter that reflects the dynamics of single-molecule passage through the nanogap and thus holds important physical and chemical implications. To clarify this, we examined potential of t_d with molecular length, molecular weight, isoelectric point, and signal frequency. As shown in Figure 2f and Figure S6, t_d does not correlate with molecular length, molecular weight, or isoelectric point, but shows a clear correlation with signal frequency. Detailed investigations of the fluid dynamics of single molecules within the nanogap are planned as part of future work.

3. *The fundamental quantum electron tunneling strategy that is frequently used by the authors should be briefly mentioned in the introduction. Besides, the authors should clearly demonstrate, which changes in this strategy are essential for successful discrimination of chiral molecules compared with other species. In other words, the novelty of the current work in terms of technical aspects should be indicated.*

<Reply3>

The primary technical innovation of this study, detailed in the Methods section and Figure S12, lies in the development of a highly stable system capable of fabricating nanogaps with sub-nanometer precision. To clarify this contribution and emphasize the novelty of our approach for chiral discrimination, we have added the following statement to the Introduction.

<Main text>

Here, we use an adjustable gap between two atomically pointed gold nanowires—formed with high stability and sub-nanometer precision using a mechanically controllable break-junction (MCBJ)—to monitor molecular identity.

4. Reviewer 1, Comment 19: PBS buffer at pH 7.4 was used. Does the type of buffer affect the discrimination performance of the method? By considering isoelectric points of amino acids, it is still not clear why pH was set at 7.4.

<Reply4>

We anticipate that single-molecule conductance would be significantly affected by large pH variations, as pH alters the electronic states of amino acids and the charges of their amino and carboxyl groups. At very low or high pH, the elevated proton or hydroxide ions concentrations can generate substantial background currents, potentially exceeding the dynamic range of our measurement system. In addition, in view of future efforts to enhance measurement throughput via electrophoresis, operating at the isoelectric point would be unsuitable due to suppressed electrophoretic mobility. Based on these considerations, we selected a standard biological buffer and used PBS at pH 7.4 for all measurements.

Reviewer #3 (Remarks to the Author):

The revised version of the manuscript is greatly improved and the authors have made many of the revisions suggested. However they still have not performed an analysis on a natural sample which I believe is essential for this publication to appear in any of the Nature journals, or any other high impact journal for that matter. (My Comment 3). It is not possible to simply "explain away" the need for verification of an analytical technique. The purpose of testing on a real sample is to test the "unknown unknowns" and it can't be replaced by laboratory simulants. To be completely transparent with the authors I will say that this has been my personal experience as a science author publishing papers on the development of new analytical techniques as well, and I think this is fair, and a legitimate test. I have never been able to publish a new technique without analyzing a real sample. This critical test shows that the system would really work in a real world scenario. Also for the record, keep in mind that this paper is intended for a journal literally named Nature. If this paper is about a new way to explore nature it should definitely examine a natural sample.

<Reply>

We thank the reviewer for this insightful suggestion and agree that evaluating the feasibility of our single-molecule analysis method on real-world samples is essential.

In response, we analyzed sediment samples from the Atacama Desert—a terrestrial analog of planetary surfaces—and a sample of the Murchison meteorite from the Smithsonian Institution using our single-molecule method in collaboration with NASA. To validate our findings, we also analyzed the same samples by mass spectrometry.

Our results show that the meteorite sample is racemic. The sand samples were nearly racemic, although some exhibited a slight excess of the L-enantiomer. These results have been incorporated into the main text, Figure 5, and the Supporting Information.

<Main text>

Quantification and identification of amino acids extracted from the Murchison meteorite and the Atacama desert

Given the promising results with various mixtures of molecules presented above, we sought to determine whether our instrument and analysis could identify and quantify amino acids in natural samples of astrobiological relevance. We therefore conducted single-molecule tunneling current measurements on amino acid extracts from the Murchison meteorite, and from sediment samples from two sites of the Atacama Desert, referred to as North of Antofagasta (NOA) and Diego de Almagro (DGA) (Supplementary Information S4). Parallel analyses were performed using liquid chromatography-mass spectrometry (LC-MS), a gold-standard for chiral amino acid analyses. Relative comparison of nanogap signal events with amino acid concentrations estimated via LC-MS revealed a strong agreement for the Murchison meteorite and the DGA samples (Fig. 5a). However, higher signal event count was detected in the NOA samples and blank sample compared to the lower amino acid concentrations quantified via LC-MS, suggesting possible contributions from background contamination during nanogap analysis, likely originating from adsorbed organic residues on the chip surface or trace airborne contaminants introduced during sample preparation (Fig. 5a).

For compound-specific analysis, six target amino acids—glycine, alanine, valine, serine, aspartic acid, and glutamate—were selected based on a previous study⁵⁰. Although many additional amino acids are known to be present in these samples, we focused on these seven species for consistency with prior work. All detected signals in the nanogap analysis were classified into the trained categories, meaning that any signal not belonging to these categories was assigned to the most similar class based on the classifier's decision boundaries, as no explicit "others" class was included. Among these, five chiral amino acids (alanine, valine, serine, aspartic acid, and glutamate) were individually classified into their D- and L-enantiomers, yielding 11 molecular species for classification. Single-molecule events were extracted from the acquired current-time profiles and classified using machine learning models trained for each amino acid. When applied to the Murchison sample, glycine was identified as the most abundant amino acid (63%), followed by glutamate (14%) and alanine (12%) (Fig. 5b, Table 3). These results closely matched LC-MS composition ratios—glycine (62.83%), glutamate (16.44%), and alanine (15.28%)—demonstrating the quantitative reliability of the nanogap single-molecule measurement method.

A similar approach was applied to the NOA and DGA samples to estimate the relative abundances of each amino acid. In the NOA samples, strong consistency with LC-MS results was

observed for glycine (NG: $65\% \pm 10\%$; LC-MS: $59.63\% \pm 7.49\%$), alanine (NG: $12\% \pm 7\%$; LC-MS: $8.43\% \pm 0.55\%$), and glutamate (NG: $14\% \pm 8\%$; LC-MS: $4.02\% \pm 1.29\%$). However, serine and valine showed considerable variability, likely due to limited classification accuracy near the detection limit (Fig. 5b, Table 3). In the DGA samples, strong agreement was again observed for glycine (NG: $56\% \pm 6\%$; LC-MS: $51.27\% \pm 1.75\%$), as well as for alanine, aspartic acid, and glutamate (Fig. 5b, Table 3). In contrast, serine and valine continued to exhibit high variation, and no consistent trend was observed for these low-abundance amino acids.

Application of chiral classification to amino acids extracted from samples of the Murchison meteorite and the Atacama Desert

Chiral discrimination of alanine, valine, serine, aspartic acid, and glutamate was performed on samples from the Murchison meteorite and the NOA and DGA sites in the Atacama Desert. D/L ratios were calculated based on the relative signal counts of the D- and L-enantiomers and compared with values obtained from LC-MS analysis. Overall, the nanogap-derived D/L ratios showed good agreement with LC-MS results (Fig. 5c, Table 4). Linear regression analysis across all samples yielded a coefficient of determination (R^2) of 0.719, indicating a significant positive correlation between the results from the two techniques. For the Murchison sample, several major amino acids, including alanine, showed strong consistency between the techniques. In contrast, the LC-MS D/L ratio for glutamate in the DGA sample was 0.13 ± 0.09 , compared to 0.70 ± 0.43 from the nanogap analysis. Although these values overlap within the experimental uncertainty, the relatively large error in the nanogap measurement likely reflects the limited number of detected signal events and the intrinsic variability of single-molecule tunneling signals. Moreover, greater variability in D/L ratios was observed for amino acids present at low concentrations. In particular, serine and valine generated few events for robust statistical evaluation, leading to reduced quantitative accuracy. Extending the measurement time did not substantially increase the number of detectable events, likely due to molecular adsorption and gradual degradation of the nanogap stability over prolonged operation.

Overall, these results demonstrate the capability of the nanogap method to quantify enantiomeric compositions with high reliability, supporting its use for D/L ratio estimation. However, for amino acids present at extremely low concentrations or in highly skewed

enantiomeric ratios, improvements in classification accuracy and system-level corrections remain necessary.

Discussion

In this study, we demonstrated the applicability of nanogap-based single-molecule detection for astrobiologically relevant samples through the analysis of amino acids extracted from extraterrestrial and terrestrial materials. Quantitative nanogap-derived profiles from the Murchison meteorite closely matched LC-MS results, confirming the accuracy of the method. In the Atacama Desert samples (NOA and DGA), the relative abundances and D/L ratios showed both consistencies and discrepancies with LC-MS data, reflecting the possible influence of background contamination and highlighting the high sensitivity of the nanogap system, especially in low-abundance regimes. These findings support the feasibility of using single-molecule tunneling measurements for quantitative and stereochemical profiling of amino acids in extraterrestrial and extreme terrestrial environments. Conventional LC-MS techniques, while highly sensitive on Earth, are often impractical for in situ astrobiological missions due to their size, vacuum and solvent requirements, and need for complex sample preparation. In contrast, the nanogap (NG) platform offers a compact, reagent-free, and potentially integrable approach for on-site molecular chirality detection, making it highly promising for future planetary exploration applications.

METHODS

Amino acids extraction from astrobiologically relevant samples. Amino acids were extracted from the Murchison meteorite and two sites in the Atacama Desert [North of Antofagasta (NOA) and Diego de Almagro (DGA)]. For each extraction, ~1.2 g of dried sample was mechanically homogenized and subjected to hot-water extraction at 100°C for 24 h, followed by centrifugation, multiple washes, and drying. The NOA and DGA extractions were performed in triplicates. A portion (45%) of each extract then underwent vapor-phase acid hydrolysis at 150°C for 3 h in 6 M HCl. Amino acids were further purified via cation-exchange chromatography and eluates were dried and stored before analysis. The resulting acid-hydrolyzed samples were reconstituted in Milli-Q ultrapure water and analyzed by both liquid chromatography-mass spectrometry (LC-MS) and single-molecule detection (Supporting Information S5, Fig. S13). For the latter, the dried samples were rehydrated in Milli-Q ultrapure water to yield amino acid concentrations between

1 and 100 μM , optimal for nanogap measurements. Due to higher total amino acid content, Murchison extracts were diluted in 10 mL of water, whereas NOA and DGA extracts were rehydrated in 1 mL. All single-molecule measurements were conducted using freshly prepared 50 μL aliquots to ensure consistency. For more details about the samples, the amino acid extraction workflow, and specifications about the LC-MS and single-molecule analyses refer to Supplementary Information 3-10, Figures S13-18, and Table S2.

Figure

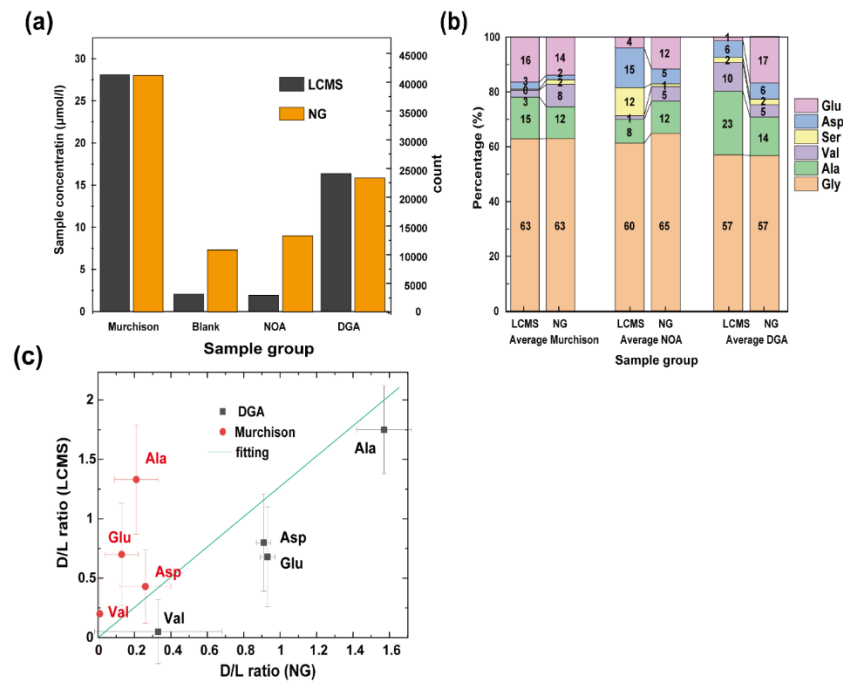


Figure 5. Quantification and comparative analysis of 6 amino acids in the Murchison meteorite and the Atacama Desert samples by LC-MS and nanogap analysis. a, Comparison of amino acid content and single-molecule detection signals across the studied samples. Black bars represent amino acid concentrations (left axis, $\mu\text{mol/L}$) as determined by LC-MS analysis, while orange bars indicate the number of single-molecule tunneling events (right axis, count) detected using the nanogap (NG) measurement method. Samples include the Murchison meteorite extract, blank control, Atacama NOA, and Atacama DGA. A general correlation between concentration and signal frequency is observed; deviations in low-concentration samples are likely attributable to background contamination. **b,** Quantification of amino acids in extracts from Murchison, Atacama NOA, and Atacama DGA. Stacked bar chart illustrating the relative composition of amino acids identified in each sample. Each bar shows the proportion of six major amino acids as determined by LC-MS and single-molecule nanogap measurements: glycine

(orange), glutamate (pink), aspartic acid (blue), serine (yellow), valine (purple), and alanine (green). The relative abundances, measured by NG, reflect the fractions of tunneling signals attributed to each species, highlighting sample-specific compositional differences. c, Correlation between D/L ratios obtained from LC-MS and NG measurements. The x-axis represents D/L ratios measured by the nanogap method, while the y-axis shows those determined by LC-MS.

Table 3. Amino acid relative abundances (%) of 6 amino acids in the 6 M HCl-hydrolyzed hot water extracts of the Murchison meteorite and the Atacama Desert samples (DGA and NOA) measured by LC-FD/Q-ToF-MS and nanogap analysis.^a

Sample	Murchison		NOA		DGA	
Method	LC-MS	NG	LC-MS	NG	LC-MS	NG
Glycine	62.83 ± 0.69	0.63 ± 0.11	59.63 ± 7.49	0.65±0.10	51.27 ± 1.75	0.56±0.06
Alanine	15.28 ± 0.36	0.12 ± 0.07	8.43 ± 0.55	0.12±0.07	20.85 ± 0.97	0.14±0.01
Valine	2.55 ± 0.72	0.08 ± 0.02	1.23 ± 0.29	0.05±0.05	9.43 ± 0.16	0.05±0.03
Serine	0.34 ± 0.01	0.02 ± 0.01	11.74 ± 5.25	0.01±0.01	1.69 ± 0.30	0.02±0.01
Aspartic Acid	2.57 ± 0.05	0.03 ± 0.02	14.95 ± 2.13	0.05±0.03	5.60 ± 1.94	0.06±0.03
Glutamate	16.44 ± 0.30	0.14 ± 0.07	4.02 ± 1.29	0.12±0.08	11.16 ± 0.86	0.17±0.09

^aValues are means of $n = 3$ independent runs. For each run r , relative abundances were computed as $p_{i,r} = \frac{A_{i,r}}{\sum_k A_{k,r}}$. The table reports $\bar{p}_i \pm SE(\bar{p}_i)$, where $\bar{p}_i$ is the across-run mean. The standard errors (SE) were computed as $\frac{s}{\sqrt{n}}$ after per-run normalization, where s denotes the sample standard deviation across runs. Percent values were obtained by multiplying $\bar{p}_i$ and $SE(\bar{p}_i)$ by 100.

Table 4. Blank-corrected D/L ratios of 5 chiral amino acids in the 6 M HCl-hydrolyzed hot-water extracts of the Murchison meteorite and the Atacama desert samples (DGA and NOA) measured by LC-FD/Q-ToF-MS and nanogap analysis.^a

Sample	Murchison		NOA		DGA	
Method	LC-MS	NG	LC-MS	NG	LC-MS	NG
Alanine	1.66±0.1 5	1.75±0.3 7	n.d.	1.22±0.40	0.21±0.12	1.33±0.46
Valine	0.35±0.3 5	0.05±0.2 7	n.d.	0.13±0.47	0.01±0.01	0.20±0.31
Serine	n.d.	0.15±0.3 2	n.d.	0.29±0.25	n.d.	0.14±0.23
Aspartic acid	0.91±0.0 4	0.80±0.4 1	n.d.	0.23±0.41	0.26±0.14	0.43±0.31
Glutamate	0.95±0.0 4	0.68±0.4 2	n.d.	0.61±0.33	0.13±0.09	0.70±0.43

n.d. = value not determined due to trace amino acid abundances.

^aD/L ratios were computed from the run-level means $\bar{D}$ and $\bar{L}$. Standard errors (SE) were obtained by

first-order error propagation assuming independent $\bar{D}$ and $\bar{L}$: $SE\left(\frac{\bar{D}}{\bar{L}}\right) \approx \frac{\bar{D}}{\bar{L}} \sqrt{\left(\frac{SE(\bar{D})}{\bar{D}}\right)^2 + \left(\frac{SE(\bar{L})}{\bar{L}}\right)^2}$,

$SE(\bar{x}) = \frac{s_x}{\sqrt{n}}$. Here s_x is the sample standard deviation across runs.

Table 5. Blank-corrected L-enantiomeric excesses (L_{ee}) of 5 chiral amino acids in the 6 M HCl-hydrolyzed hot water extracts of the Murchison meteorite and the Atacama desert samples (DGA and NOA) measured by LC-FD/Q-ToF-MS and nanogap analysis.^a

Sample	Murchison		NOA		DGA	
Method	LC-MS	NG	LC-MS	NG	LC-MS	NG
Alanine	-24.73±0.04	-27.39±0.82	n.d.	8.54±42.32	65.82±0.17	35.01±23.31
Valine	47.89±0.39	89.67±62.87	n.d.	79.36±83.62	97.40±0.03	91.81±78.33
Serine	n.d.	74.34±22.86	n.d.	50.83±23.82	n.d.	19.53±54.54
Aspartic acid	4.59±0.02	10.81±0.79	n.d.	62.66±73.07	59.35±0.17	60.17±73.79
Glutamate	2.64±0.02	18.66±1.61	n.d.	-11.46±46.03	76.40±0.10	69.77±42.40

n.d. = value not determined due to trace amino acid abundances.

aLee was defined as $ee_L(\%) = \frac{100(\bar{L}-\bar{D})}{(\bar{L}+\bar{D})}$. Standard errors were obtained by first-order error propagation from $SE(\bar{L}) = s_L/\sqrt{n_L}$ and $SE(\bar{D}) = s_D/\sqrt{n_D}$ (assuming independence): $SE(ee_L\%) = 100 \cdot \frac{2}{(\bar{L}+\bar{D})^2} \sqrt{\bar{D}^2 SE(\bar{L})^2 + \bar{L}^2 SE(\bar{D})^2}$. Here s_L and s_D are sample standard deviations across runs.

<Supporting Information>

S3. 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 MΩ·cm, < 3 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 pg/mL) to minimize metal contamination during amino acid hydrolysis.. Cation-exchange resin (AG50W-X8, 100-200 mesh, NH_4OH were 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).

S4. Samples collection.

Atacama Desert samples were collected during a field expedition in February 2022. Two sampling sites were targeted: Diego de Almagro (DGA; 26°23'11" S, 70°04'14" W) and North of Antofagasta (NOA; 22°47'46" S, 70°04'14" 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.

S5. 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 (Figure S14). After heating, the ampoules were opened and centrifuged at 3000 rpm for three minutes, the water supernatants were then transferred from the ampoules by pipetting into pre-weighed amber glass vials (Figure S15). 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 (Figure S16). 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 (Figure S17).

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 (Figure S18).

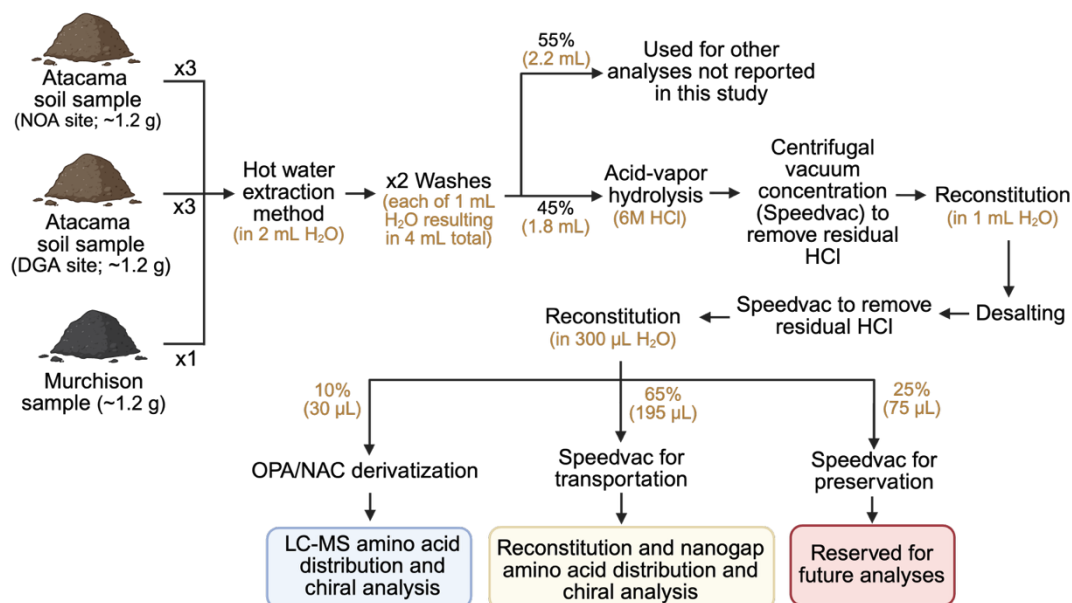


Figure S1. Summary of amino acids extraction procedure.

Schematic summary of the extraction workflow applied to the Murchison meteorite fragment and the two Atacama Desert soils (NOA and DGA sites). The figure shows sample preparation, extraction, acid-hydrolysis, desalting, reconstitutions, and aliquoting, with volumes and percentages allocates LC-MS, nanogap analysis, and future analyses.

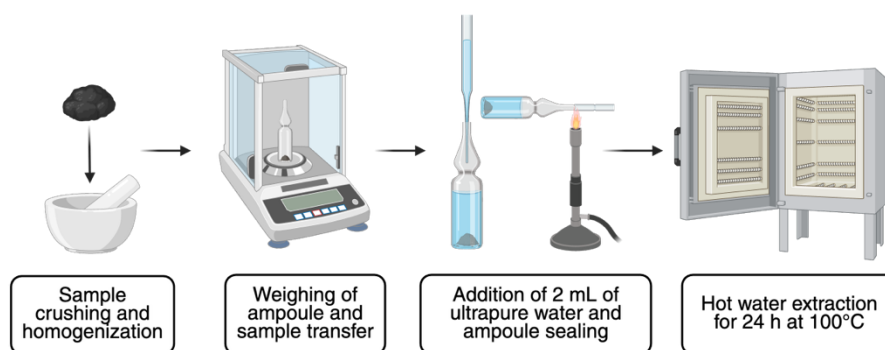


Figure S2. Sample preparation and extraction workflow.

Illustration of the procedure for sample preparation and hot-water extraction from Atacama Desert soils and the Murchison meteorite sample.

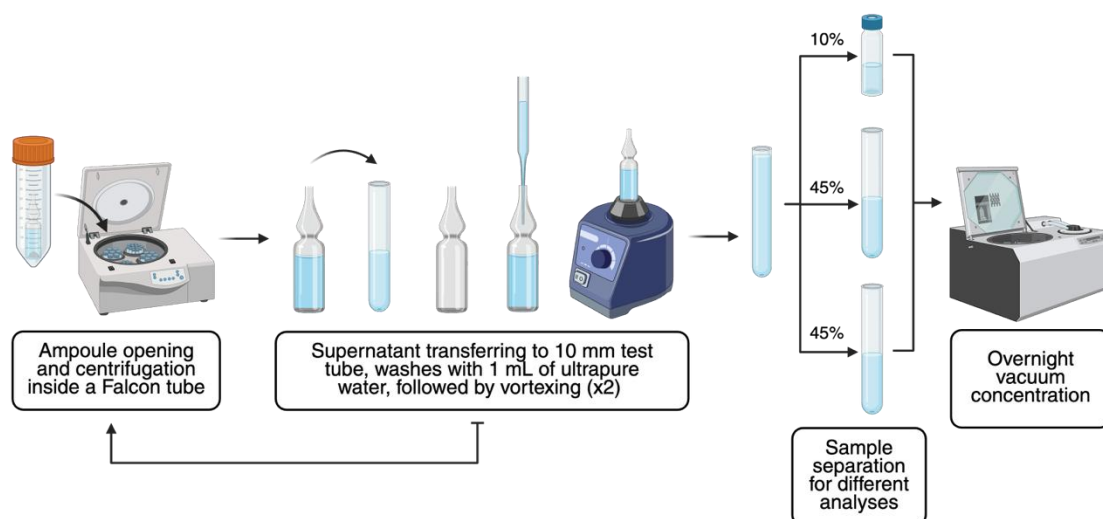


Figure S3. Washes and dry-down workflow.

Schematic of the washes and dry-down procedure, showing sample splitting for non-hydrolyzed analysis (45%), hydrolyzed analysis (45%), and additional analyses (10%) not reported in this study.

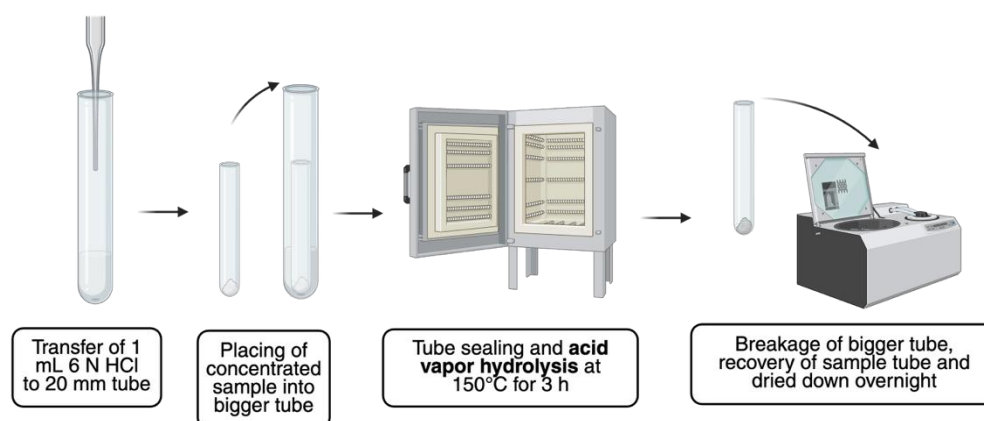


Figure S4. Acid-hydrolysis workflow.

Schematic showing the vapor-phase acid-hydrolysis performed on the hot-water extracts from the Atacama Desert soils and the Murchison meteorite sample.

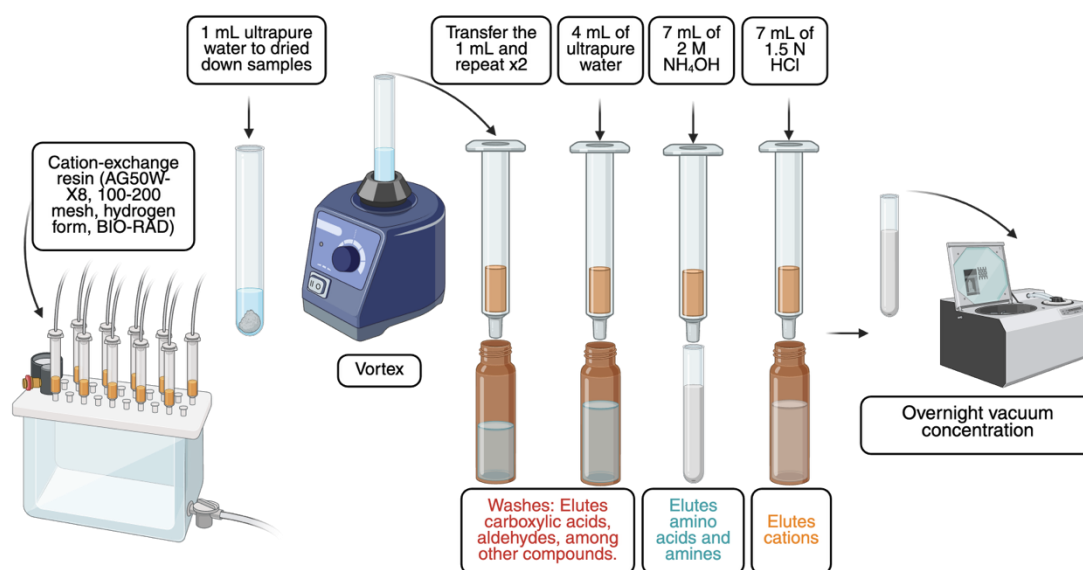


Figure S5. Amino acid purification workflow.

Schematic of the cation exchange purification procedure. Bulk organic extracts from environmental samples were fractionated into three eluates: carboxylic acids and aldehydes, amino acids and amines, and cations. Only the amino acid/amine fraction was analyzed and reported here.

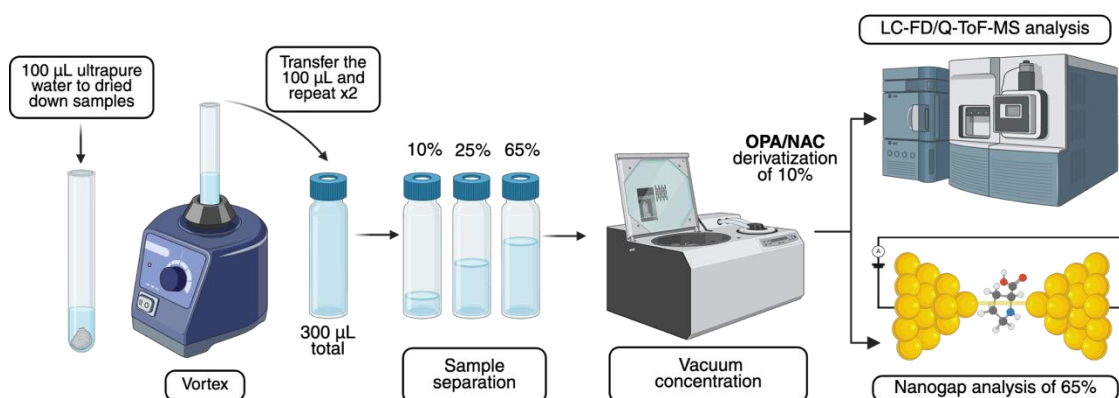


Figure S6. Sample preparation for LC-MS and nanogap analysis.

Schematic of the wash steps applied to desalted dry-down samples, followed by separation for LC-MS and nanogap analyses.

S6. 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 (Table S2), the concentration of the amino acids quantified in this sample was estimated at 28.1 µM. 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 micropercent (µ%)—**corresponding to molecules that diffused into and transiently occupied the nanogap volume (~femtoliter scale)—**were detected as signal events. This fraction reflects **the limited number of molecules physically entering the tunneling junction during the measurement time**, rather than incomplete ergodicity of the solution. This extremely low detection rate is consistent with the intrinsic limitations of single-molecule detection in 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 ($\sim 10\ \mu\text{L}$) was required to maintain stable electrical contact and prevent electrode drying during long-term measurements.

S7. Evaluation of amino acid composition in natural samples from the Murchison meteorite and Atacama Desert via single-molecule analysis

The amino acid composition of natural samples from the Murchison meteorite and the NOA and DGA sites in the Atacama Desert was assessed using single-molecule detection and LC-MS analysis. A previous study (Amashukeli et al., 2007) identified seven amino acids—glutamate, aspartic acid, serine, valine, alanine, proline, and glycine—in Atacama soil samples using high-performance liquid chromatography (HPLC). Based on this precedent, the present study focused on six of these amino acids for single-molecule and LC-MS. Due to the derivatization agent used for the LC-MS workflow, it was not possible to quantify D- and L-proline, and therefore, these were also excluded from single-molecule analysis. Among the remaining six, five (glutamate, aspartic acid, serine, valine, and alanine) are chiral, having both D- and L-enantiomers. These enantiomers were treated as distinct classification targets, resulting in 11 molecular species (including glycine). Identification was performed by extracting single-molecule events from the acquired current–time profile data and training machine learning–based classification models for each amino acid species. More than 9,000 signal events per amino acid were used for training, and classification accuracy was evaluated using the F-measure.

When this classifier was applied to the acid-hydrolyzed Murchison sample, which yielded a total of 41,158 signal events, glycine was identified as the most frequently detected amino acid, accounting for 62.9% (25,898 signals) of all events. The next most abundant were glutamate (5,711 signals; 13.9%) and alanine (4,801 signals; 11.7%; Fig. 5b and Table 3). The same sample was also analyzed by LC-MS, and the resulting amino acid composition was compared with that

obtained by single-molecule analysis. The relative abundances measured by LC-MS were 62.83% for glycine, 16.44% for glutamate, and 15.28% for alanine—values closely aligned with the single-molecule detection results. Valine also showed good agreement between the two methods, with a relative abundance of 4.4% from nanogap measurements and 2.55% from LC-MS. Similar trends were observed for serine, and aspartic acid, further supporting the consistency and quantitative reliability of the nanogap-based single-molecule sensing approach relative to the conventional LC-MS analysis.

S8. Relative quantification of amino acids in Atacama NOA and DGA samples using single-molecule detection

In addition to the Murchison sample, single-molecule signals obtained from the Atacama NOA and DGA acid-hydrolyzed samples were analyzed using the nanogap (NG) electrode method to determine the relative proportions of individual amino acids. The resulting ratios were compared with corresponding values obtained from LC-MS analysis.

In the NOA samples, glycine was the most frequently detected amino acid, with a signal ratio of $65\% \pm 10\%$, closely matching the LC-MS-determined ratio of $59.63\% \pm 7.49\%$. Alanine (LC-MS: $8.43\% \pm 0.55\%$; NG: $12\% \pm 7\%$) and glutamate (LC-MS: $4.02\% \pm 1.29\%$; NG: $12\% \pm 8\%$) also showed reasonable agreement between the two methods. Aspartic acid exhibited somewhat lower concordance (LC-MS: $14.95\% \pm 2.13\%$; NG: $5\% \pm 3\%$), though the values were still within an acceptable range, supporting the utility of nanogap-based relative quantification for these amino acids. In contrast, serine and valine showed greater variability— $1\% \pm 1\%$ and $6\% \pm 5\%$, respectively—compared to their LC-MS counterparts ($11.74\% \pm 5.25\%$ for serine and $1.23\% \pm 0.29\%$ for valine). This discrepancy likely reflects challenges in identifying low-abundance species. In the case of serine, the absolute concentration may have approached the detection limit, limiting classification accuracy. Similarly, low valine concentrations likely contributed to the increased variability.

In the DGA samples, glycine (LC-MS: $51.27\% \pm 1.75\%$; NG: $56\% \pm 6\%$), alanine (LC-MS: $20.85\% \pm 0.97\%$; NG: $14\% \pm 1\%$), aspartic acid (LC-MS: $5.60\% \pm 1.94\%$; NG: $6\% \pm 3\%$), and glutamate (LC-MS: $11.16\% \pm 0.86\%$; NG: $17\% \pm 9\%$) showed generally good agreement between nanogap- and LC-MS-based quantification. Glycine was the most abundant amino acid in the DGA samples, yielding 12,432 signal events—second only to the Murchison sample—indicating high quantitative reliability. However, as in the NOA samples, serine and valine exhibited substantial

variability in the DGA samples, with nanogap-derived ratios of $2\% \pm 1\%$ and $5\% \pm 3\%$, respectively. These discrepancies likely reflect concentrations near the detection threshold and the intrinsic challenges of minority class identification in classification models.

Taken together, these findings demonstrate that, as with the Murchison sample, relative quantification of amino acids in the NOA and DGA samples using the nanogap method is feasible. Notably, glycine, alanine, aspartic acid, and glutamate showed strong concordance with LC-MS-based results. Conversely, for low-abundance amino acids such as serine and valine, further improvements in identification accuracy and measurement sensitivity remain necessary.

S9. Chiral isomer discrimination and D/L ratio evaluation in natural samples from the Murchison meteorite, and two sites of the Atacama Desert (NOA and DGA)

Chiral isomer discrimination of amino acids in natural samples derived from the Murchison meteorite and the Atacama Desert NOA and DGA sites was conducted using LC-MS and nanogap, and D/L ratios were evaluated. The analysis focused on five amino acids—alanine, valine, serine, aspartic acid, and glutamate—for which both L- and D-enantiomers were individually identified. D/L ratios were calculated from the relative signal counts corresponding to each enantiomer and compared with values obtained by LC-MS (Fig. 5c and Table 4).

Overall, the D/L ratios obtained via the nanogap method showed good agreement with those measured by LC-MS. A linear regression analysis across all samples yielded a coefficient of determination (R^2) of 0.719, indicating a significant positive correlation between the two methods. This result suggests that nanogap-based single-molecule detection can reliably capture the quantitative trends of chiral isomer distributions.

In the Murchison sample, the D/L ratios of major amino acids, such as alanine, aspartic acid, and glutamate were highly consistent between the two methods. Notably, the D/L ratio for alanine was 1.66 ± 0.15 by LC-MS and 1.75 ± 0.37 by nanogap detection, indicating strong concordance. In contrast, the D/L ratio measured by LC-MS for glutamate in the DGA sample was substantially lower (0.13 ± 0.19) than that determined by nanogap analysis (0.70 ± 0.43), reflecting a marked discrepancy. This may result from underestimation of D-enantiomer content in the LC-MS analysis or from systematic error in chiral discrimination within the nanogap platform.

A general trend was also observed in which D/L ratio variability increased as amino acid concentration decreased. For example, in the Murchison sample, nanogap-derived D/L ratios for serine and valine were 0.15 ± 0.32 and 0.05 ± 0.27 , respectively, reflecting substantial uncertainty and limited quantitative accuracy. These deviations likely stem from the low absolute abundance of these amino acids, leading to insufficient signal events for robust statistical evaluation. This concentration-dependent increase in D/L ratio variability suggests that stochastic detection of fewer signal events at low concentrations leads to greater statistical uncertainty in chirality estimation. A similar trend was also observed in LC–MS measurements, where low-abundance species exhibited larger relative errors in D/L quantification, consistent with signal-to-noise limitations inherent to both methods.

In summary, the nanogap-based method provides a reliable means for estimating D/L ratios of chiral isomers in naturally derived amino acids. However, for low-abundance species or samples with extreme enantiomeric skew, further improvements in discrimination accuracy and implementation of system-level correction strategies remain necessary.

S10. LC-FD/ToF-MS analyses.

Amino acid abundances, distribution, and enantiomeric ratios were determined by LC-FD/ToF-MS using the methodology previously reported in Glavin et al. 2020. The desalted 10% hydrolyzed and non-hydrolyzed samples set aside for OPA/NAC analysis were dried down overnight for LC-FD/ToF-MS analysis. From this reconstituted sample, 20 μ L was dried down with 20 μ L of sodium borate buffer (pH = 9). After dry down, the samples were brought up in 20 μ L of Milli-Q ultrapure water and 5 μ L of 0.1 M OPA/NAC derivatization agent and allowed to react for 15 minutes at room temperature before quenching with 75 μ L of 0.1 M hydrazine. The amino acids in the NH_4OH eluates were derivatized with OPA/NAC for 15 minutes at room temperature followed by their separation and analysis using a Waters ACQUITY UPLC and Waters Xevo G2-XS Q-ToF-MS operating in positive ion mode. C_2 to C_6 amino acids were chromatographically resolved using a Waters BEH C18 column (2.1 $\times$ 50 mm, 1.7 μ m bead) and a Waters BEH phenyl column (2.1 $\times$ 150 mm, 1.7 μ m bead) in series. Both columns were maintained at 30.0°C. The mobile phase conditions for amino acid separations were as follows: flow rate, 150 μ L/min; gradient, time in minutes (%B): 0 (0), 35 (55), 45 (100). C_5 amino acid isomers and enantiomers were chromatographically separated using the same chromatography conditions as for the C_2 to

C₆ amino acids but required the implementation of a different gradient. The gradient used for C₅ amino acid isomers and enantiomers was structured via time in minutes (%B): 0 (15), 25 (20), 25.06 (35), 44.5 (40), 45 (100).

During the Xevo G2-XS analysis, a dual electrospray ionization (ESI) system was used for the purpose of implementing lock mass corrections. The primary ESI source was operated using the following parameters: capillary voltage, 3.0 kV; sampling cone voltage, 40 V; source temperature, 120°C; desolvation gas (N₂) temperature, 350°C; cone gas (N₂) flow, 50 L h⁻¹, desolvation gas flow rate, 750 L h⁻¹. Due to the possibility that minor variations in the mass-to-charge (*m/z*) scale may occur while executing experimental runs after instrument calibration is performed, a reference ESI source was implemented to supply an independent leucine enkephalin standard signal. The reference ESI source was operated using a sample infusion rate of 20 µL min⁻¹, a sample fill volume of 250 µL, a lockspray infusion rate of 10 µL min⁻¹, a capillary voltage of 3.0 kV, a reference cone voltage of 30 V, and a collision energy of 6.0 V. The ToF analyzer was operated in “Sensitivity mode,” which used a reflectron to provide a full width at half maximum resolution of <22,000 based on the [M+H]⁺ of leucine enkephalin, *m/z* 556.2771.

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, λ_{ex} = 340 nm, λ_{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.

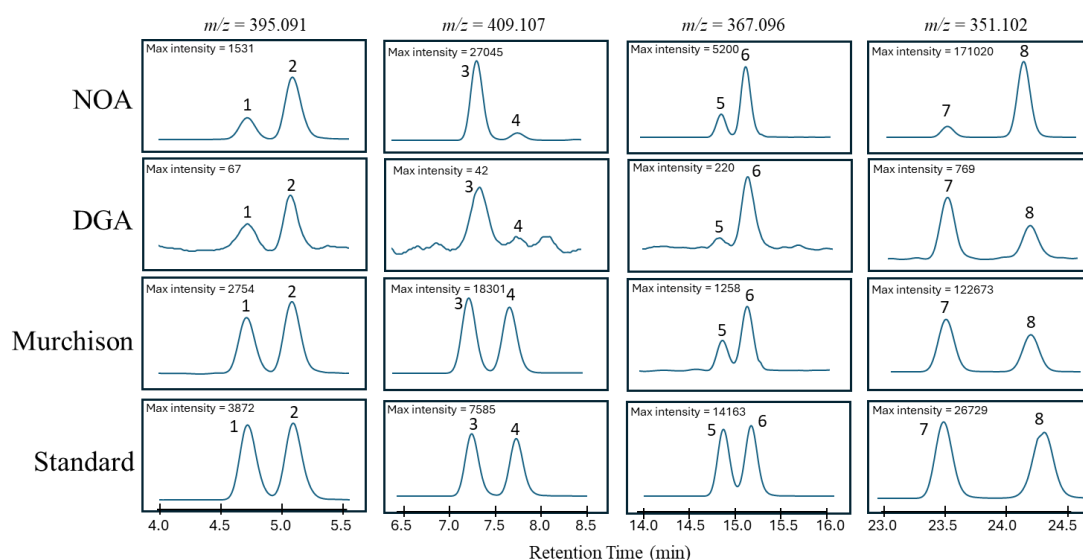


Figure S7. LC-MS-ToF chromatograms of Asp, Glu, Ser, and Ala.

Chromatograms show elution of aspartic acid ($m/z = 395.0913$), glutamate ($m/z = 409.1069$), serine ($m/z = 367.0964$), and alanine ($m/z = 351.1015$) in the acid-hydrolyzed, hot-water extracts of Murchison meteorite, Atacama DGA sample, Atacama NOA sample and the standard. The 4- to 6-min region of the $m/z = 395.0913$, the 6.5- to 8.5-min region of the $m/z = 409.1069$, the 14- to 16-min region of the $m/z = 367.0964$, and the 23- to 24.5-min of the $m/z = 351.1015$ of the LC-ToF selected-ion chromatograms corresponding to the *o*-phthaldialdehyde/*N*-acetyl-L-cysteine (OPA/NAC) derivatives in positive ion mode via heated electrospray ionization and a 5 ppm mass accuracy. Similar single-ion chromatograms were obtained for the non-hydrolyzed water extracts. Peaks were identified by comparisons of their retention times and exact monoisotopic masses with those in the amino acid standard analyzed on the same day, and are designated by peak number as follows: (1) D-aspartic acid, (2) L-aspartic acid, (3) L-glutamate, (4) D-glutamate, (5) D-serine, (6) L-serine, (7) D-alanine, (8) L-alanine.

Table S1. 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 this 28 species.

S11. 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 (Fig. S20). The four tested alcohols (Figure. S21) 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 (Figure S20). For machine learning,

90% of each dataset was used for training and the remaining 10% for testing. Discrimination accuracy, evaluated using the F1-score, reached 96.2% for the (S)/(R) pair of 1-phenylbutan-1-ol (Fig. S21b) and 97.0% for 1-(naphthalen-1-yl)ethan-1-ol (Fig. S21c), 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.

Response to Reviewers

We sincerely thank the Editor and the reviewers for their careful evaluation of our manuscript and for their constructive feedback. We have carefully considered all comments and have revised the manuscript accordingly. Below, we provide point-by-point responses to each issue raised. All changes to the manuscript are indicated in bold or as tracked changes.

Reviewer #1 General Comment:

Reviewer #1 (Remarks to the Author): I have thoroughly reviewed the manuscript, along with the author's responses to feedback from other reviewers. I am pleased to note that the manuscript has undergone substantial improvements, and the concerns I previously had have been effectively addressed in the revised draft. Given these improvements, I wholeheartedly recommend this article for publication in its current form.

<Reply>

We are grateful to Reviewer #1 for their positive assessment and recommendation of this manuscript. We have further strengthened the manuscript in response to the remaining comments from Reviewer #3, as detailed below.

Reviewer #3 General Comments:

Reviewer #3 (Remarks to the Author): I appreciate the authors' substantial efforts to address my prior concern regarding validation on natural samples. The inclusion of analyses of the Murchison meteorite and Atacama Desert samples greatly improves this manuscript. However, after reviewing the revised version carefully, several important issues remain that I believe require clarification prior to publication.

<Reply>

We are grateful to Reviewer #3 for their thorough and critical assessment of our work. Their comments have significantly improved the clarity and rigor of the manuscript. We have addressed each concern as detailed below.

Reviewer Comment: Issue 1. Forced Classification Without an “Unknown” Category

The manuscript states that “All detected signals in the nanogap analysis were classified into the trained categories... as no explicit ‘others’ class was included.” This raises a significant analytical concern. In complex environmental samples, signals originating from contaminants or untrained molecular species will necessarily be assigned to the closest trained amino acid category. This forced-classification framework can inflate apparent agreement with LC-MS and distort both compositional and D/L ratio estimates, particularly at low abundance.

I strongly recommend that the authors do these things or equivalent to clarify the situation:

- A) Clarify how forced classification may bias quantitative estimates.*
- B) Report classifier confidence metrics or probability thresholds.*
- C) Discuss how implementation of an “unknown” rejection class would affect interpretation.*
- D) Provide blank classification composition data to show how background signals are distributed across classes.*

Without this clarification, the robustness of the method in heterogeneous real-world samples remains uncertain.

<Reply>

We thank the reviewer for this important comment. We agree that the original closed-set analysis had an important limitation: events originating from process-related background signals, contaminants, or molecular species not represented in the reference panel could be assigned to the nearest trained amino-acid class. This may bias estimates of target-panel composition and D/L ratios, particularly in low-abundance samples and complex environmental matrices.

The environmental-sample analysis was designed as a targeted analysis rather than a comprehensive profiling of all amino acids or organic constituents. The target panel comprised glycine and the D- and L-enantiomers of alanine, valine, serine, aspartic acid, and glutamic acid, yielding 11 predefined molecular species. These targets were selected a priori based on previous reports of amino acids in related Atacama Desert soil samples and their compatibility with the LC–MS workflow. We have revised the manuscript to clarify that the resulting values represent relative abundances within this predefined target panel and do not represent the absolute molecular composition of the environmental extracts.

A) Potential bias arising from forced classification.

We now explicitly state that the original closed-set analysis could overestimate the apparent abundance of target species if nontarget events were assigned to the nearest target class. This may affect the relative composition of the six-amino-acid groups and the corresponding D/L ratios. Accordingly, we no longer interpret the original closed-set outputs as direct quantitative estimates of all molecular species present in the extracts.

B) Confidence-based rejection analysis.

To address this issue, we re-analyzed the environmental and process-blank datasets using the same probability-based rejection criterion. For each feature-extracted event, the maximum predicted probability among the 11 target classes was calculated. Events with a maximum class probability ≥ 0.5 were retained as high-confidence target events, whereas events with a maximum probability < 0.5 were classified as unknown/rejected events and excluded from the revised compositional and D/L-ratio analyses. The 0.5 threshold was used as a predefined operational criterion to distinguish events with a dominant target-class assignment from lower-confidence events; its impact on event composition is reported transparently in the revised analysis.

The resulting target fractions, defined as the proportion of high-confidence target events relative to all feature-extracted events, were 49.7% for Murchison, 28.3% for DGA, 25.4% for NOA, and 6.8% for the process blank. Thus, 93.2% of feature-extracted blank events were rejected under the same criterion. These results indicate that most process-background events were not retained as high-confidence assignments to the targeted amino-acid panel.

C) Interpretation of the unknown/rejected category.

The unknown/rejected category does not represent a chemically identified “other” molecular class. It includes events that did not meet the predefined confidence criterion for assignment to any of the 11 target species. Such events may include process-derived background signals, nontarget molecules in the environmental matrix, low-quality events, or target-derived events

located near a classification boundary. We have revised the manuscript to clearly distinguish high-confidence target assignments from low-confidence events and to clarify that the method is currently a targeted identification approach rather than an open-set molecular profiling method.

D) Blank classification composition.

We additionally report the class composition of the process blank. Before confidence filtering, blank events were distributed across all 11 trained classes, consistent with the reviewer's concern that closed-set classification can distribute background events among target categories. After applying the probability threshold, only 727 of 10,681 blank feature-extracted events remained as high-confidence target assignments. The retained blank assignments were not uniformly distributed, with the following six-analyte compositions: glycine, 33.7%; alanine, 5.9%; valine, 7.7%; serine, 0.4%; aspartic acid, 51.3%; and glutamic acid, 1.0%. These results are provided in Supplementary Tables S4 and S5.

The revised analyses shown in Figs. 5b and 5c were calculated using only high-confidence target events. This reanalysis retained a substantial high-confidence target fraction in the Murchison sample, whereas DGA and NOA showed lower target fractions. We now interpret NOA more cautiously as a low-abundance and matrix-sensitive sample. We have revised the Results, Discussion, figure legends, and Methods accordingly and added the full event-flow analysis, from total picked-up events to feature-extracted events, high-confidence target events, and unknown/rejected events, in Supplementary Tables S5 and S6.

Revised Results, "Identification and quantification of amino acids extracted from samples of the Murchison meteorite and the Atacama Desert," pp. 17–20; revised Figs. 5b and 5c; revised Tables 3 and 4; revised Methods, "Machine learning," pp. 26–27; Supplementary Tables S5 and S6.

Results (pp. 17–18)

Identification and quantification of amino acids extracted from the Murchison meteorite and Atacama Desert samples

Given the promising results obtained with various molecular mixtures described above, we sought to determine whether our instrument and analysis pipeline could identify and quantify amino acids in natural samples of astrobiological relevance. We therefore conducted single-molecule tunneling-current measurements on amino-acid extracts from the Murchison meteorite and soil samples from two sites in the Atacama Desert, referred to as North of Antofagasta (NOA) and Diego de Almagro (DGA) (Supplementary Information S4). In parallel, we

performed liquid chromatography–mass spectrometry (LC–MS), which serves as a reference method for chiral amino-acid characterization.

Figure 5a shows a comparison of the total amino acid abundance determined via LC–MS with the total number of tunneling-current events detected after signal acquisition via nanogap measurements. Total nanogap event counts include both events subsequently retained as high-confidence target assignments and events that did not meet the confidence criterion. Therefore, total event counts should not be interpreted as direct quantitative measures of total amino acid abundance. Murchison and DGA showed higher total event counts than the process blank, whereas NOA showed event counts closer to the blank level, indicating a larger relative contribution of process-derived background signals and/or nontarget matrix components under low-abundance conditions.

For compound-specific analysis, we performed a targeted analysis using glycine and the D- and L-enantiomers of alanine, valine, serine, aspartic acid, and glutamic acid. These analytes were selected a priori based on previous Atacama soil studies and compatibility with the LC–MS workflow. Glycine was treated as a single achiral species, whereas the remaining five amino acids were resolved into D- and L-enantiomers, yielding an 11-species target panel.

The environmental-sample analysis was designed as targeted relative quantification within this predefined panel rather than comprehensive profiling of all amino acids or other organic constituents in the extracts. To reduce the forced assignment of low-confidence events, the maximum predicted probability across the 11 target classes was calculated for each feature-extracted event. Events with a maximum probability ≥ 0.5 were retained as high-confidence target events, whereas events below this threshold were classified as unknown/rejected events and excluded from the revised compositional and D/L-ratio analyses.

The proportions of high-confidence target events among feature-extracted events were 49.7% for Murchison, 28.3% for DGA, 25.4% for NOA, and 6.8% for the process blank. Accordingly, 93.2% of feature-extracted blank events were classified as unknown/rejected under the same criteria. These results show that the confidence-based filter removed most blank-derived events from the targeted environmental-sample analysis.

The relative compositions of the six-amino-acid groups were recalculated using only retained high-confidence target events. For each chiral amino acid, D- and L-enantiomer counts were combined, and the resulting six-group counts were normalized to the total number of retained high-confidence target events.

In the Murchison sample, glycine remained the most abundant group in the confidence-filtered nanogap analysis (47.9%), followed by alanine (16.9%), glutamic acid (12.0%), and serine (11.7%) (Fig. 5b and Table 3). This composition retained the major glycine-, alanine-, and glutamic acid-rich patterns observed using LC–MS, although analyte-specific differences remained between the two methods, including higher nanogap-derived proportions of serine and valine compared to those detected via LC–MS.

In the NOA and DGA samples, confidence filtering retained sample-specific target panel compositions but revealed analyte-dependent deviations from LC–MS results. In NOA, glycine, alanine, serine, and aspartic acid showed broadly comparable relative abundances between the two methods, whereas valine was higher and glutamic acid was not retained as a high-confidence nanogap assignment. In DGA, glycine and alanine showed similar relative contributions between the two methods, whereas valine, serine, and aspartic acid showed higher nanogap-derived proportions, and glutamic acid was not retained after confidence filtering. Thus, the confidence-filtered analysis supports targeted compositional comparison while indicating that quantitative agreement depends on both analyte identity and sample matrix.

Application of chiral classification to amino acids extracted from the Murchison meteorite and Atacama Desert samples

Chiral discrimination of alanine, valine, serine, aspartic acid, and glutamic acid was performed for amino-acid extracts from the Murchison meteorite and the NOA and DGA sites in the Atacama Desert. For the confidence-filtered analysis, D/L ratios were calculated using only retained high-confidence D- and L-enantiomer assignments and were compared with the corresponding values obtained via LC–MS.

Figure 5c shows a comparison of D/L ratios of alanine, valine, and aspartic acid in the Murchison and DGA extracts, for which ratios could be determined using both methods. Glutamic acid was excluded because its confidence-filtered nanogap-derived D/L ratio could not be obtained for the DGA sample. Serine was also excluded because the number of retained high-confidence assignments was insufficient to determine reliable D/L ratios under the confidence-filtered analysis.

The comparison showed analyte- and sample-dependent agreement rather than uniform quantitative correspondence across all amino acids. For the Murchison sample, the confidence-filtered nanogap-derived D/L ratio for aspartic acid was in good agreement with the corresponding LC–MS value, whereas differences remained for alanine and valine. The DGA sample likewise showed analyte-dependent discrepancies between the two methods. These results indicate that confidence filtering reduces contributions from ambiguous events while highlighting the current limitations of quantitative chiral analysis in chemically complex environmental matrices.

Accordingly, the confidence-filtered nanogap D/L ratios should be interpreted as high-confidence estimates within the predefined targeted amino-acid panel rather than as comprehensive chiral profiles of the environmental samples. Further validation using matrix-matched reference materials, increased event statistics, and the implementation of open-set recognition approaches will be required to establish practical quantitative performance for chemically heterogeneous environmental samples.

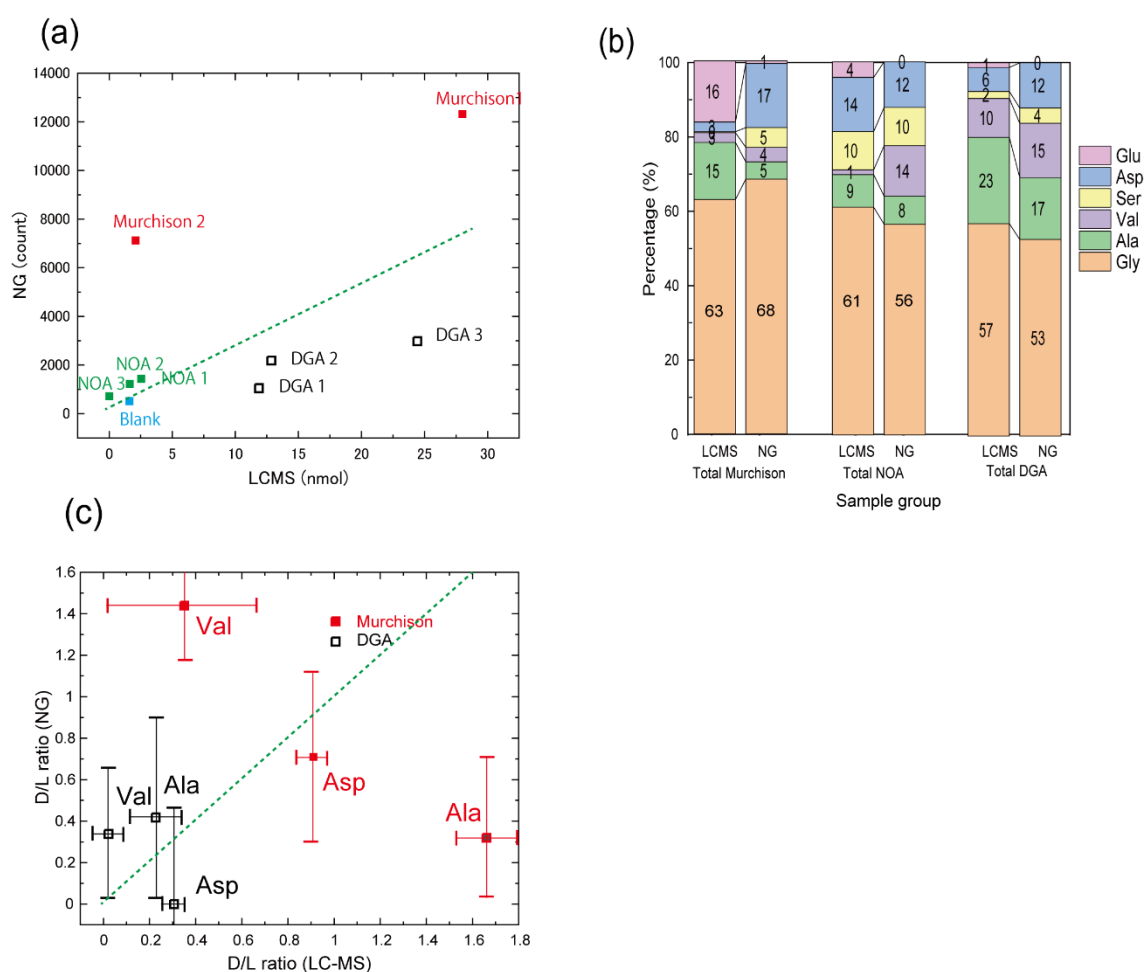


Figure 5 captions:

Figure 5 Confidence-filtered targeted analysis of amino acids in extracts from Murchison and Atacama Desert. **a**, Relationship between the total amino-acid abundance determined by LC–MS and the total tunnelling-current event counts detected after signal pickup by nanogap (NG) measurements for the Murchison meteorite extract, process blank, and Atacama Desert NOA and DGA extracts. Each point represents one independently prepared extract analysed by both LC–MS and NG measurements (Murchison, $n = 2$; NOA, $n = 3$; DGA, $n = 3$; Blank, $n = 1$). Total NG event counts include both target-like events and non-target or low-confidence events. The dashed line represents the least-squares linear regression. **b**, Relative abundances of glycine, alanine, valine, serine, aspartic acid, and glutamic acid determined via LC–MS and confidence-filtered NG analysis for the Murchison, NOA, and DGA extracts. NG values were calculated using only high-confidence events assigned to a predefined 11-species target panel comprising glycine and the D- and L-enantiomers of the five chiral amino acids. Events with a maximum class probability < 0.5 were classified as unknown/rejected and excluded from the analysis. D- and L-enantiomer counts were combined for each chiral amino acid, and the resulting six-group

abundances were normalized to the total number of retained high-confidence target events. NG-derived values represent the relative abundances within the predefined target panel rather than the overall molecular composition of the extracts. Error bars indicate mean $\pm$ s.e.m. from three independent NG measurements. **c**, Comparison of D/L ratios determined via LC–MS and confidence-filtered NG analysis for alanine, valine, and aspartic acid in the Murchison and DGA extracts. The x-axis represents LC–MS D/L ratios, and the y-axis represents NG D/L ratios calculated from retained high-confidence assignments. Squares indicate the Murchison sample, and circles indicate the DGA sample. Error bars indicate the mean $\pm$ s.e.m. from three independent measurements. The dashed line indicates the line of identity ($y = x$).

Reviewer Comment: Issue 2. Background Contamination and Elevated Blank Signals

The authors note elevated signal event counts in NOA and blank samples relative to LC-MS quantification and attribute this to surface or airborne contamination.

Given the high sensitivity of the nanogap platform, contamination control becomes a huge issue in terms of quantitative reliability.

Please clarify whether background subtraction was performed and whether contamination produces systematic enrichment in particular amino acid classes. For a technique proposed for planetary exploration, contamination susceptibility must be carefully quantified.

<Reply>

We thank the reviewer for emphasizing the critical importance of contamination control, particularly given the high sensitivity of the nanogap platform. We agree that the elevated event counts in low-abundance environmental samples and the process blank require careful interpretation and that the susceptibility of the method to contamination must be explicitly addressed, especially in the context of future planetary exploration applications.

In the original analysis, no formal background subtraction was applied to the nanogap event counts before classification or compositional analysis. We have revised the manuscript to clarify this point. Rather than performing a numerical blank subtraction, we re-analyzed the environmental and process blank datasets using the same confidence-based rejection procedure described in our response to Issue 1.

For each feature-extracted event, the maximum predicted probability across the predefined 11-species target panel was calculated. Events with a maximum probability < 0.5 were classified as unknown/rejected and excluded from the revised targeted compositional and D/L-ratio analyses.

This confidence-based filtering substantially reduced the contribution of process-derived background events while avoiding assumptions inherent in direct blank subtraction. The revised analysis therefore provides a more conservative and robust assessment of targeted amino-acid composition and chirality in complex environmental samples.

[Revised Results, “Identification and quantification of amino acids extracted from samples of the Murchison meteorite and the Atacama Desert,” pp. 16–17; revised Methods, “Machine learning,” pp. 25–26.]

Identification and quantification of amino acids extracted from the Murchison meteorite and Atacama Desert samples

Given the promising results obtained with various molecular mixtures described above, we sought to determine whether our instrument and analysis pipeline could identify and quantify amino acids in natural samples of astrobiological relevance. We therefore conducted single-molecule tunneling-current measurements on amino-acid extracts from the Murchison meteorite and soil samples from two sites in the Atacama Desert, referred to as North of Antofagasta (NOA) and Diego de Almagro (DGA) (Supplementary Information S4). In parallel, we performed liquid chromatography–mass spectrometry (LC–MS), which serves as a reference method for chiral amino-acid characterization.

Figure 5a shows a comparison of the total amino acid abundance determined via LC–MS with the total number of tunneling-current events detected after signal acquisition via nanogap measurements. Total nanogap event counts include both events subsequently retained as high-confidence target assignments and events that did not meet the confidence criterion. Therefore, total event counts should not be interpreted as direct quantitative measures of total amino acid abundance. Murchison and DGA showed higher total event counts than the process blank, whereas NOA showed event counts closer to the blank level, indicating a larger relative contribution of process-derived background signals and/or nontarget matrix components under low-abundance conditions.

For compound-specific analysis, we performed a targeted analysis using glycine and the D- and L-enantiomers of alanine, valine, serine, aspartic acid, and glutamic acid. These analytes were selected a priori based on previous Atacama soil studies and compatibility with the LC–MS workflow. Glycine was treated as a single achiral species, whereas the remaining five amino acids were resolved into D- and L-enantiomers, yielding an 11-species target panel.

The environmental-sample analysis was designed as targeted relative quantification within this predefined panel rather than comprehensive profiling of all amino acids or other organic constituents in the extracts. To reduce the forced assignment of low-confidence events, the maximum predicted probability across the 11 target classes was calculated for each feature-extracted event. Events with a maximum probability ≥ 0.5 were retained as high-confidence

target events, whereas events below this threshold were classified as unknown/rejected events and excluded from the revised compositional and D/L-ratio analyses.

The proportions of high-confidence target events among feature-extracted events were 49.7% for Murchison, 28.3% for DGA, 25.4% for NOA, and 6.8% for the process blank. Accordingly, 93.2% of feature-extracted blank events were classified as unknown/rejected under the same criteria. These results show that the confidence-based filter removed most blank-derived events from the targeted environmental-sample analysis.

The relative compositions of the six-amino-acid groups were recalculated using only retained high-confidence target events. For each chiral amino acid, D- and L-enantiomer counts were combined, and the resulting six-group counts were normalized to the total number of retained high-confidence target events.

In the Murchison sample, glycine remained the most abundant group in the confidence-filtered nanogap analysis (47.9%), followed by alanine (16.9%), glutamic acid (12.0%), and serine (11.7%) (Fig. 5b and Table 3). This composition retained the major glycine-, alanine-, and glutamic acid-rich patterns observed using LC-MS, although analyte-specific differences remained between the two methods, including higher nanogap-derived proportions of serine and valine compared to those detected via LC-MS.

In the NOA and DGA samples, confidence filtering retained sample-specific target panel compositions but revealed analyte-dependent deviations from LC-MS results. In NOA, glycine, alanine, serine, and aspartic acid showed broadly comparable relative abundances between the two methods, whereas valine was higher and glutamic acid was not retained as a high-confidence nanogap assignment. In DGA, glycine and alanine showed similar relative contributions between the two methods, whereas valine, serine, and aspartic acid showed higher nanogap-derived proportions, and glutamic acid was not retained after confidence filtering. Thus, the confidence-filtered analysis supports targeted compositional comparison while indicating that quantitative agreement depends on both analyte identity and sample matrix.

Methods pp. 25–26

Machine learning. Machine-learning analysis was performed using Python 3.9.7, scikit-learn version 0.24.2, and the XGBoost classifier. Pulse signals extracted from the current-time profiles were converted into 15-dimensional feature vectors comprising the maximum current, the current amplitude relative to the baseline, the event duration, and 12 time-resolved current features. The latter features were calculated as the average current within 12 equal-duration intervals spanning each event.

For the classification of amino-acid standards, random undersampling was applied to reduce class imbalance. The number of feature vectors for each molecular species was reduced to match that of the smallest class. The resulting dataset was partitioned into o subsets and

evaluated using 10-fold cross-validation, in which one subset was used for testing and the remaining nine subsets were used for training. Classification performance was reported as the mean accuracy obtained across the 10 rounds of classification.

For environmental-sample analysis, a predefined 11-species target panel was used, comprising glycine and the D- and L-enantiomers of alanine, valine, serine, aspartic acid, and glutamic acid. Events for which all 15 features could be extracted were defined as feature-extracted events. For each feature-extracted event, the trained XGBoost classifier generated predicted probabilities for all 11 target classes. The highest predicted probability among the 11 classes was defined as the event-level classification confidence.

For the confidence-filtered environmental analysis, events with a maximum predicted probability ≥ 0.5 were retained as high-confidence target events. Events with a maximum predicted probability < 0.5 were classified as unknown/rejected and excluded from the targeted compositional and D/L-ratio analyses. The same probability threshold was applied to the Murchison, NOA, DGA, and process blank datasets without sample-specific adjustment. The unknown/rejected category does not represent a chemically identified molecular class; rather, it includes events that did not meet the predefined confidence criterion for assignment to any of the 11 target species.

The total number of picked-up events was used for the comparison shown in Fig. 5a. The compositional analysis shown in Fig. 5b and Table 3 and the D/L-ratio analysis shown in Fig. 5c and Table 4 were calculated using only retained high-confidence target events. For each independent nanogap measurement, D- and L-enantiomer counts were combined to determine the relative abundance of each chiral amino acid, and the resulting abundances were normalized to the total number of retained high-confidence target events. Relative abundances and D/L ratios are reported as the mean $\pm$ s.e.m. from three independent nanogap measurements.

This analysis showed that only 727 of 10,681 feature-extracted blank events (6.8%) in the process blank were retained as high-confidence target assignments, whereas the remaining 93.2% were classified as unknown/rejected. In comparison, the proportions of high-confidence target events were 49.7% for the Murchison sample, 28.3% for DGA, and 25.4% for NOA. These results indicate that the majority of process-derived background events did not satisfy the predefined confidence criterion for assignment to the targeted amino-acid panel.

[Revised Results, pp. 17–18; Supplementary Table S4.]

Identification and quantification of amino acids extracted from the Murchison meteorite and Atacama Desert samples

Given the promising results obtained with various molecular mixtures described above, we sought to determine whether our instrument and analysis pipeline could identify and quantify amino acids in natural samples of astrobiological relevance. We therefore conducted single-molecule tunneling-current measurements on amino-acid extracts from the Murchison meteorite and soil samples from two sites in the Atacama Desert, referred to as North of Antofagasta (NOA) and Diego de Almagro (DGA) (Supplementary Information S4). In parallel, we performed liquid chromatography–mass spectrometry (LC–MS), which serves as a reference method for chiral amino-acid characterization.

Figure 5a shows a comparison of the total amino acid abundance determined via LC–MS with the total number of tunneling-current events detected after signal acquisition via nanogap measurements. Total nanogap event counts include both events subsequently retained as high-confidence target assignments and events that did not meet the confidence criterion. Therefore, total event counts should not be interpreted as direct quantitative measures of total amino acid abundance. Murchison and DGA showed higher total event counts than the process blank, whereas NOA showed event counts closer to the blank level, indicating a larger relative contribution of process-derived background signals and/or nontarget matrix components under low-abundance conditions.

For compound-specific analysis, we performed a targeted analysis using glycine and the D- and L-enantiomers of alanine, valine, serine, aspartic acid, and glutamic acid. These analytes were selected a priori based on previous Atacama soil studies and compatibility with the LC–MS workflow. Glycine was treated as a single achiral species, whereas the remaining five amino acids were resolved into D- and L-enantiomers, yielding an 11-species target panel.

The environmental-sample analysis was designed as targeted relative quantification within this predefined panel rather than comprehensive profiling of all amino acids or other organic constituents in the extracts. To reduce the forced assignment of low-confidence events, the maximum predicted probability across the 11 target classes was calculated for each feature-extracted event. Events with a maximum probability ≥ 0.5 were retained as high-confidence target events, whereas events below this threshold were classified as unknown/rejected events and excluded from the revised compositional and D/L-ratio analyses.

The proportions of high-confidence target events among feature-extracted events were 49.7% for Murchison, 28.3% for DGA, 25.4% for NOA, and 6.8% for the process blank. Accordingly, 93.2% of feature-extracted blank events were classified as unknown/rejected under the same criteria. These results show that the confidence-based filter removed most blank-derived events from the targeted environmental-sample analysis.

The relative compositions of the six-amino-acid groups were recalculated using only retained high-confidence target events. For each chiral amino acid, D- and L-enantiomer counts were combined, and the resulting six-group counts were normalized to the total number of retained high-confidence target events.

In the Murchison sample, glycine remained the most abundant group in the confidence-filtered nanogap analysis (47.9%), followed by alanine (16.9%), glutamic acid (12.0%), and serine (11.7%) (Fig. 5b and Table 3). This composition retained the major glycine-, alanine-, and glutamic acid-rich patterns observed using LC-MS, although analyte-specific differences remained between the two methods, including higher nanogap-derived proportions of serine and valine compared to those detected via LC-MS.

In the NOA and DGA samples, confidence filtering retained sample-specific target panel compositions but revealed analyte-dependent deviations from LC-MS results. In NOA, glycine, alanine, serine, and aspartic acid showed broadly comparable relative abundances between the two methods, whereas valine was higher and glutamic acid was not retained as a high-confidence nanogap assignment. In DGA, glycine and alanine showed similar relative contributions between the two methods, whereas valine, serine, and aspartic acid showed higher nanogap-derived proportions, and glutamic acid was not retained after confidence filtering. Thus, the confidence-filtered analysis supports targeted compositional comparison while indicating that quantitative agreement depends on both analyte identity and sample matrix.

Application of chiral classification to amino acids extracted from the Murchison meteorite and Atacama Desert samples

Chiral discrimination of alanine, valine, serine, aspartic acid, and glutamic acid was performed for amino-acid extracts from the Murchison meteorite and the NOA and DGA sites in the Atacama Desert. For the confidence-filtered analysis, D/L ratios were calculated using only retained high-confidence D- and L-enantiomer assignments and were compared with the corresponding values obtained via LC-MS.

Figure 5c shows a comparison of D/L ratios of alanine, valine, and aspartic acid in the Murchison and DGA extracts, for which ratios could be determined using both methods. Glutamic acid was excluded because its confidence-filtered nanogap-derived D/L ratio could not be obtained for the DGA sample. Serine was also excluded because the number of retained high-confidence assignments was insufficient to determine reliable D/L ratios under the confidence-filtered analysis.

The comparison showed analyte- and sample-dependent agreement rather than uniform quantitative correspondence across all amino acids. For the Murchison sample, the confidence-filtered nanogap-derived D/L ratio for aspartic acid was in good agreement with the corresponding LC-MS value, whereas differences remained for alanine and valine. The DGA sample likewise showed analyte-dependent discrepancies between the two methods. These results indicate that confidence filtering reduces contributions from ambiguous events while highlighting the current limitations of quantitative chiral analysis in chemically complex environmental matrices.

Accordingly, the confidence-filtered nanogap D/L ratios should be interpreted as high-confidence estimates within the predefined targeted amino-acid panel rather than as comprehensive chiral profiles of the environmental samples. Further validation using matrix-matched reference materials, increased event statistics, and the implementation of open-set recognition approaches will be required to establish practical quantitative performance for chemically heterogeneous environmental samples.

Supplementary Table S4. 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 $\times 100$.

We additionally analyzed the composition of the retained high-confidence blank assignments. The blank distribution was nonuniform across target classes. After aggregation of D- and L-enantiomer assignments, the retained blank events comprised glycine (33.7%), alanine (5.9%), valine (7.7%), serine (0.4%), aspartic acid (51.3%), and glutamic acid (1.0%). Thus, although confidence filtering substantially reduced the contribution of blank-derived events, the remaining blank assignments showed class-dependent enrichment, particularly for aspartic acid. We therefore revised the manuscript to avoid interpreting the confidence-filtered nanogap composition as an absolute amino-acid composition of the environmental extracts. Instead, the results are presented strictly as relative compositions within the retained high-confidence target panel.

[Revised Results, pp. 17–18; revised Fig. 5b caption; revised Table 3; Supplementary Table S5.]

Identification and quantification of amino acids extracted from the Murchison meteorite and Atacama Desert samples

Given the promising results obtained with various molecular mixtures described above, we sought to determine whether our instrument and analysis pipeline could identify and quantify amino acids in natural samples of astrobiological relevance. We therefore conducted single-molecule tunneling-current measurements on amino-acid extracts from the Murchison meteorite and soil samples from two sites in the Atacama Desert, referred to as North of Antofagasta (NOA) and Diego de Almagro (DGA) (Supplementary Information S4). In parallel, we performed liquid chromatography–mass spectrometry (LC–MS), which serves as a reference method for chiral amino-acid characterization.

Figure 5a shows a comparison of the total amino acid abundance determined via LC–MS with the total number of tunneling-current events detected after signal acquisition via nanogap measurements. Total nanogap event counts include both events subsequently retained as high-confidence target assignments and events that did not meet the confidence criterion. Therefore, total event counts should not be interpreted as direct quantitative measures of total amino acid abundance. Murchison and DGA showed higher total event counts than the process blank, whereas NOA showed event counts closer to the blank level, indicating a larger relative contribution of process-derived background signals and/or nontarget matrix components under low-abundance conditions.

For compound-specific analysis, we performed a targeted analysis using glycine and the D- and L-enantiomers of alanine, valine, serine, aspartic acid, and glutamic acid. These analytes were selected a priori based on previous Atacama soil studies and compatibility with the LC–MS workflow. Glycine was treated as a single achiral species, whereas the remaining five amino acids were resolved into D- and L-enantiomers, yielding an 11-species target panel.

The environmental-sample analysis was designed as targeted relative quantification within this predefined panel rather than comprehensive profiling of all amino acids or other organic constituents in the extracts. To reduce the forced assignment of low-confidence events, the maximum predicted probability across the 11 target classes was calculated for each feature-extracted event. Events with a maximum probability ≥ 0.5 were retained as high-confidence target events, whereas events below this threshold were classified as unknown/rejected events and excluded from the revised compositional and D/L-ratio analyses.

The proportions of high-confidence target events among feature-extracted events were 49.7% for Murchison, 28.3% for DGA, 25.4% for NOA, and 6.8% for the process blank. Accordingly, 93.2% of feature-extracted blank events were classified as unknown/rejected

under the same criteria. These results show that the confidence-based filter removed most blank-derived events from the targeted environmental-sample analysis.

The relative compositions of the six-amino-acid groups were recalculated using only retained high-confidence target events. For each chiral amino acid, D- and L-enantiomer counts were combined, and the resulting six-group counts were normalized to the total number of retained high-confidence target events.

In the Murchison sample, glycine remained the most abundant group in the confidence-filtered nanogap analysis (47.9%), followed by alanine (16.9%), glutamic acid (12.0%), and serine (11.7%) (Fig. 5b and Table 3). This composition retained the major glycine-, alanine-, and glutamic acid-rich patterns observed using LC-MS, although analyte-specific differences remained between the two methods, including higher nanogap-derived proportions of serine and valine compared to those detected via LC-MS.

In the NOA and DGA samples, confidence filtering retained sample-specific target panel compositions but revealed analyte-dependent deviations from LC-MS results. In NOA, glycine, alanine, serine, and aspartic acid showed broadly comparable relative abundances between the two methods, whereas valine was higher and glutamic acid was not retained as a high-confidence nanogap assignment. In DGA, glycine and alanine showed similar relative contributions between the two methods, whereas valine, serine, and aspartic acid showed higher nanogap-derived proportions, and glutamic acid was not retained after confidence filtering. Thus, the confidence-filtered analysis supports targeted compositional comparison while indicating that quantitative agreement depends on both analyte identity and sample matrix.

Application of chiral classification to amino acids extracted from the Murchison meteorite and Atacama Desert samples

Chiral discrimination of alanine, valine, serine, aspartic acid, and glutamic acid was performed for amino-acid extracts from the Murchison meteorite and the NOA and DGA sites in the Atacama Desert. For the confidence-filtered analysis, D/L ratios were calculated using only retained high-confidence D- and L-enantiomer assignments and were compared with the corresponding values obtained via LC-MS.

Figure 5c shows a comparison of D/L ratios of alanine, valine, and aspartic acid in the Murchison and DGA extracts, for which ratios could be determined using both methods. Glutamic acid was excluded because its confidence-filtered nanogap-derived D/L ratio could not be obtained for the DGA sample. Serine was also excluded because the number of retained high-confidence assignments was insufficient to determine reliable D/L ratios under the confidence-filtered analysis.

The comparison showed analyte- and sample-dependent agreement rather than uniform quantitative correspondence across all amino acids. For the Murchison sample, the confidence-

filtered nanogap-derived D/L ratio for aspartic acid was in good agreement with the corresponding LC–MS value, whereas differences remained for alanine and valine. The DGA sample likewise showed analyte-dependent discrepancies between the two methods. These results indicate that confidence filtering reduces contributions from ambiguous events while highlighting the current limitations of quantitative chiral analysis in chemically complex environmental matrices.

Accordingly, the confidence-filtered nanogap D/L ratios should be interpreted as high-confidence estimates within the predefined targeted amino-acid panel rather than as comprehensive chiral profiles of the environmental samples. Further validation using matrix-matched reference materials, increased event statistics, and the implementation of open-set recognition approaches will be required to establish practical quantitative performance for chemically heterogeneous environmental samples.

Fig. 5b caption

Figure 5 b, Relative abundances of glycine, alanine, valine, serine, aspartic acid, and glutamic acid determined via LC–MS and confidence-filtered NG analysis for the Murchison, NOA, and DGA extracts. NG values were calculated using only high-confidence events assigned to a predefined 11-species target panel comprising glycine and the D- and L-enantiomers of the five chiral amino acids. Events with a maximum class probability < 0.5 were classified as unknown/rejected and excluded from the analysis. D- and L-enantiomer counts were combined for each chiral amino acid, and the resulting six-group abundances were normalized to the total number of retained high-confidence target events. NG-derived values represent the relative abundances within the predefined target panel rather than the overall molecular composition of the extracts. Error bars indicate mean $\pm$ s.e.m. from three independent NG measurements.

Table 3. Amino acid relative abundances (%) of six amino acids in 6 M HCl-hydrolyzed hot-water extracts of the Murchison meteorite and Atacama Desert samples (DGA and NOA) measured via LC-FD–Q-ToF-MS and nanogap analysis.^a

Sample	Murchison		NOA		DGA	
Method	LC–MS	NG	LC–MS	NG	LC–MS	NG
Glycine	62.83 $\pm$ 0.69	47.9 $\pm$ 11.0	59.63 $\pm$ 7.49	56.4 $\pm$ 10.0	51.27 $\pm$ 1.75	52.6 $\pm$ 6.0

Alanine	15.28 ± 0.36	16.9 ± 7.0	8.43 ± 0.55	7.5 ± 7.0	20.85 ± 0.97	16.6 ± 1.0
Valine	2.55 ± 0.72	8.2 ± 2.0	1.23 ± 0.29	13.5 ± 5.0	9.43 ± 0.16	14.5 ± 3.0
Serine	0.34 ± 0.01	11.7 ± 1.0	11.74 ± 5.25	10.3 ± 1.0	1.69 ± 0.30	4.1 ± 1.0
Aspartic Acid	2.57 ± 0.05	3.3 ± 2.0	14.95 ± 2.13	12.2 ± 3.0	5.60 ± 1.94	12.1 ± 3.0
Glutamic acid	16.44 ± 0.30	12.0 ± 7.0	4.02 ± 1.29	0.0 ± 0.0	11.16 ± 0.86	0.0 ± 0.0

^aValues are means of $n = 3$ independent runs. For each run r , relative abundances were computed as $p_{i,r} = \frac{A_{i,r}}{\sum_k A_{k,r}}$. The table reports $\underline{p}_i \pm SE(\underline{p}_i)$, where $\underline{p}_i$ is the across-run mean. Standard errors (SEs) were computed as $\frac{s}{\sqrt{n}}$ after per-run normalization, where s denotes the sample standard deviation across runs. Percent values are reported as $\underline{p}_i \times 100$ and $SE(\underline{p}_i) \times 100$.

Supplementary Table S5. 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

Molecular species	Event count	Fraction of retained high-confidence blank events (%)
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%.

These blank data demonstrate that contamination and/or nontarget matrix components can contribute nonuniformly to retained target assignments. This issue is particularly relevant for low-abundance samples, such as NOA, and for analytes whose nanogap-derived abundances overlap with blank-associated classes. Accordingly, we have revised the Results and Discussion sections to interpret the NOA and DGA datasets more cautiously and to highlight the need for matrix-matched calibration, contamination-controlled sample preparation, representative nontarget reference datasets, and open-set recognition approaches.

[Revised Results, pp. 18–19; revised Discussion, pp. 20–21; revised LOD discussion, pp. 8–9.]

Discussion (pp 20–21)

Herein, we evaluated nanogap-based single-molecule analysis using amino-acid extracts from astrobiologically relevant samples. Confidence-based filtering enabled separation of high-confidence target assignments from low-confidence events within a predefined 11-species amino-acid panel. The process blank showed a low fraction of high-confidence target assignments, whereas the Murchison sample exhibited a substantially higher retained target fraction. The DGA and NOA samples showed intermediate target fractions, indicating that process-derived background signals and matrix-dependent non-target events contribute more strongly to the results obtained for some environmental extracts.

The confidence-filtered compositional and D/L ratio analyses captured selected major features observed via LC–MS, particularly for the Murchison, and showed analyte- and sample-dependent discrepancies. The present method should therefore be interpreted as a targeted, high-confidence identification approach rather than as comprehensive molecular profiling or absolute quantification of amino-acid mixtures in environmental samples. Future work should incorporate representative blank-derived and nontarget molecular datasets, matrix-matched

calibration experiments, and open-set recognition approaches to improve discrimination between target analytes, background signals, and chemically unrepresented events.

Discussion (pp 8–9)

A target detection limit of ~1 nM has been proposed for life-detection measurements in planetary exploration¹⁶, and reported amino-acid abundances in some meteorites exceed this concentration¹⁷. The calibration-based LOD obtained here indicates that the nanogap system is sensitive under controlled measurement conditions. However, this value should not be interpreted as the practical identification or quantification limit in chemically complex environmental matrices, where process-derived background signals, nontarget molecular species, and confidence-based event rejection can substantially affect the number of retained target assignments. Establishing practical performance in such samples will require matrix-matched calibration and contamination-controlled validation experiments.

For future planetary applications, contamination control will require rigorous chip surface preparation, high-purity reagent and solvents, procedural blanks, and repeated blank measurements throughout the analytical workflow. We have additionally included event-flow analysis and the class composition of blank assignments before and after confidence filtering in Supplementary Fig. 4 and Supplementary Table S5.

[Revised Discussion, pp. 20–21; Supplementary Tables S4 and S5]

Supplementary Table S4. 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 S5. 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%.

Reviewer Comment: Issue 3. D/L Ratio Correlation Strength

The reported correlation between nanogap-derived and LC-MS D/L ratios yields $R^2 = 0.719$. This level of agreement is encouraging and suggests that the method is capable of recovering meaningful chiral information from natural samples. However, the correlation is still moderate, and the relatively large uncertainties observed for several amino acids (particularly those at low abundance) indicate that enantiomeric ratio determination remains sensitive to event statistics and classification uncertainty.

To better evaluate the quantitative performance of the method, it would be helpful if the authors provide additional statistical detail such as: regression slope and intercept with confidence intervals, statistical tests of deviation from unity slope, or if they give a discussion of potential systematic bias, if present. Because chiral discrimination is a central claim of the manuscript, providing this additional analysis would strengthen the quantitative interpretation of the results. But again to be clear these are very encouraging results that should be strengthened for a Nature publication.

<Reply>

We thank the reviewer for this constructive suggestion and agree that the quantitative interpretation of nanogap-derived D/L ratios requires careful statistical evaluation.

In the original manuscript, we reported a regression analysis between nanogap-derived and LC-MS D/L ratios, including an overall coefficient of determination of $R^2 = 0.719$. However, following the confidence-based reanalysis described in our responses to Issues 1 and 2, we concluded that this regression is no longer an appropriate summary of quantitative performance. The original analysis relied on forced closed-set assignments, whereas the revised analysis excludes feature-extracted events with a maximum predicted probability < 0.5 and restricts interpretation to retained high-confidence assignments within the predefined 11-species target panel.

After confidence filtering, directly comparable D/L ratios were available only for alanine, valine, and aspartic acid in the Murchison and DGA extracts. We therefore removed the regression line, regression statistics, and associated claims of an overall correlation from Fig. 5c and the main text. The revised figure instead shows the directly comparable D/L ratios together with the line of identity, allowing the level of agreement and the deviations between the nanogap and LC-MS measurements to be assessed without implying a general linear correlation or calibration between the two methods.

The revised analysis shows analyte- and sample-dependent agreement rather than uniform quantitative correspondence. For the Murchison sample, the confidence-filtered nanogap-derived D/L ratio for aspartic acid was comparable to the LC-MS value, whereas larger discrepancies remained for alanine and valine. The DGA sample likewise showed analyte-

dependent deviations between the two methods. We now explicitly state that confidence-filtered D/L ratios should be interpreted as targeted, high-confidence estimates within the predefined amino-acid panel rather than as comprehensive quantitative chiral profiles of complex environmental extracts.

The relatively large uncertainties associated with several D/L ratios are consistent with the limited numbers of retained high-confidence events and the inherent uncertainty in class classification, particularly for low-abundance or matrix-rich samples. We have therefore expanded the Discussion to emphasize the need for increased event statistics, matrix-matched validation, and open-set recognition approaches before practical quantitative performance can be established for heterogeneous environmental samples.

[Revised Results, “Application of chiral classification to amino acids extracted from the Murchison meteorite and Atacama Desert samples,” pp. 18–19; revised Fig. 5c and Table 4; revised Discussion, pp. 20–21.]

Application of chiral classification to amino acids extracted from the Murchison meteorite and Atacama Desert samples

Chiral discrimination of alanine, valine, serine, aspartic acid, and glutamic acid was performed for amino-acid extracts from the Murchison meteorite and the NOA and DGA sites in the Atacama Desert. For the confidence-filtered analysis, D/L ratios were calculated using only retained high-confidence D- and L-enantiomer assignments and were compared with the corresponding values obtained via LC–MS.

Figure 5c shows a comparison of D/L ratios of alanine, valine, and aspartic acid in the Murchison and DGA extracts, for which ratios could be determined using both methods. Glutamic acid was excluded because its confidence-filtered nanogap-derived D/L ratio could not be obtained for the DGA sample. Serine was also excluded because the number of retained high-confidence assignments was insufficient to determine reliable D/L ratios under the confidence-filtered analysis.

The comparison showed analyte- and sample-dependent agreement rather than uniform quantitative correspondence across all amino acids. For the Murchison sample, the confidence-filtered nanogap-derived D/L ratio for aspartic acid was in good agreement with the corresponding LC–MS value, whereas differences remained for alanine and valine. The DGA sample likewise showed analyte-dependent discrepancies between the two methods. These results indicate that confidence filtering reduces contributions from ambiguous events while highlighting the current limitations of quantitative chiral analysis in chemically complex environmental matrices.

Accordingly, the confidence-filtered nanogap D/L ratios should be interpreted as high-confidence estimates within the predefined targeted amino-acid panel rather than as

comprehensive chiral profiles of the environmental samples. Further validation using matrix-matched reference materials, increased event statistics, and the implementation of open-set recognition approaches will be required to establish practical quantitative performance for chemically heterogeneous environmental samples.

Figure 5c

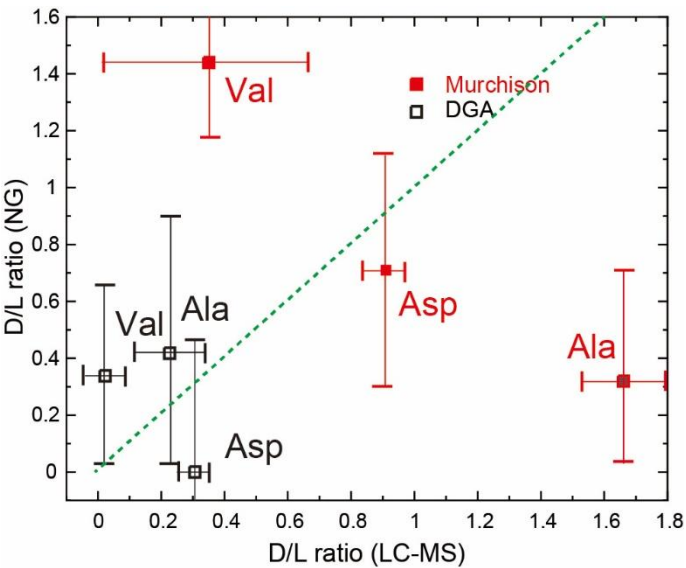


Figure 5c, Comparison of D/L ratios determined via LC–MS and confidence-filtered NG analysis for alanine, valine, and aspartic acid in the Murchison and DGA extracts. The x-axis represents LC–MS D/L ratios, and the y-axis represents NG D/L ratios calculated from retained high-confidence assignments. Squares indicate the Murchison sample, and circles indicate the DGA sample. Error bars indicate the mean $\pm$ s.e.m. from three independent measurements. The dashed line indicates the line of identity ($y = x$).

Table 4. Confidence-filtered D/L ratios of five chiral amino acids in 6 M HCl-hydrolyzed hot-water extracts of the Murchison meteorite and Atacama Desert samples measured via LC-FD–Q-ToF-MS and nanogap analysis^a.

Sample	Murchison		NOA		DGA	
Method	LC–MS	NG	LC–MS	NG	LC–MS	NG

Alanine	1.66 ± 0.15	0.32 ± 0.37	n.d.	0.47 ± 0.40	0.23 ± 0.12	0.42 ± 0.46
Valine	0.35 ± 0.35	1.44 ± 0.27	n.d.	1.22 ± 0.47	0.02 ± 0.01	0.34 ± 0.31
Serine	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Aspartic acid	0.91 ± 0.04	0.71 ± 0.41	n.d.	0.04 ± 0.41	0.31 ± 0.12	0.01 ± 0.46
Glutamic acid	0.95 ± 0.04	0.10 ± 0.42	n.d.	n.d.	0.14 ± 0.01	n.d.

^a Nanogap D/L ratios were calculated from retained high-confidence D- and L-enantiomer assignments within the predefined 11-species target panel. Events with a maximum predicted probability < 0.5 were classified as unknown/rejected and excluded from the analysis. Values are reported as mean ± s.e.m. from three independent nanogap measurements. n.d., not determined because a reliable D/L ratio could not be calculated from the retained high-confidence events.

Discussion (pp 20–21)

Herein, we evaluated nanogap-based single-molecule analysis using amino-acid extracts from astrobiologically relevant samples. Confidence-based filtering enabled separation of high-confidence target assignments from low-confidence events within a predefined 11-species amino-acid panel. The process blank showed a low fraction of high-confidence target assignments, whereas the Murchison sample exhibited a substantially higher retained target fraction. The DGA and NOA samples showed intermediate target fractions, indicating that process-derived background signals and matrix-dependent non-target events contribute more strongly to the results obtained for some environmental extracts.

The confidence-filtered compositional and D/L ratio analyses captured selected major features observed via LC–MS, particularly for the Murchison, and showed analyte- and sample-dependent discrepancies. The present method should therefore be interpreted as a targeted, high-confidence identification approach rather than as comprehensive molecular profiling or absolute quantification of amino-acid mixtures in environmental samples. Future work should incorporate representative blank-derived and nontarget molecular datasets, matrix-matched calibration experiments, and open-set recognition approaches to improve discrimination between target analytes, background signals, and chemically unrepresented events.

Methods pp. 25–26

Machine learning. Machine-learning analysis was performed using Python 3.9.7, scikit-learn version 0.24.2, and the XGBoost classifier. Pulse signals extracted from the current–time profiles were converted into 15-dimensional feature vectors comprising the maximum current, the current amplitude relative to the baseline, the event duration, and 12 time-resolved current features. The latter features were calculated as the average current within 12 equal-duration intervals spanning each event.

For the classification of amino-acid standards, random undersampling was applied to reduce class imbalance. The number of feature vectors for each molecular species was reduced to match that of the smallest class. The resulting dataset was partitioned into 10 subsets and evaluated using 10-fold cross-validation, in which one subset was used for testing and the remaining nine subsets were used for training. Classification performance was reported as the mean accuracy obtained across the 10 rounds of classification.

For environmental-sample analysis, a predefined 11-species target panel was used, comprising glycine and the D- and L-enantiomers of alanine, valine, serine, aspartic acid, and glutamic acid. Events for which all 15 features could be extracted were defined as feature-extracted events. For each feature-extracted event, the trained XGBoost classifier generated predicted probabilities for all 11 target classes. The highest predicted probability among the 11 classes was defined as the event-level classification confidence.

For the confidence-filtered environmental analysis, events with a maximum predicted probability ≥ 0.5 were retained as high-confidence target events. Events with a maximum predicted probability < 0.5 were classified as unknown/rejected and excluded from the targeted compositional and D/L-ratio analyses. The same probability threshold was applied to the Murchison, NOA, DGA, and process blank datasets without sample-specific adjustment. The unknown/rejected category does not represent a chemically identified molecular class; rather, it includes events that did not meet the predefined confidence criterion for assignment to any of the 11 target species.

The total number of picked-up events was used for the comparison shown in Fig. 5a. The compositional analysis shown in Fig. 5b and Table 3 and the D/L-ratio analysis shown in Fig. 5c and Table 4 were calculated using only retained high-confidence target events. For each independent nanogap measurement, D- and L-enantiomer counts were combined to determine the relative abundance of each chiral amino acid, and the resulting abundances were normalized to the total number of retained high-confidence target events. Relative abundances and D/L ratios are reported as the mean $\pm$ s.e.m. from three independent nanogap measurements.

Reviewer Comment: Issue 4. Au–Thiol Interaction and Dwell Time Interpretation

In response to another reviewers comment about cysteine, the authors argue that self-assembled monolayer formation requires approximately one day, and therefore Au–thiol interactions are unlikely to influence the 30-40 ms dwell time. However, this explanation may not fully address the relevant physical timescales. Au-thiol chemisorption can occur on sub-millisecond timescales and does not require any monolayer formation. The relevant question is whether transient coordination events measurably alter dwell-time distributions. But this whole topic is quite confusing because the authors actually say in the paper that they coat the gold with silica! What is the actual situation here?

Please clarify

A) whether the nanogap apex is exposed gold or coated with SiO₂,

B) Whether td distributions for cysteine show long-tail behavior indicative of transient adsorption and if the gold is not coated,

C) why transient Au–S coordination does not influence dwell time in this system.

This point is important because the manuscript attributes chiral discrimination to molecule–electrode electronic coupling. This entire discussion calls into question the overall mechanistic interpretation in the paper. The manuscript proposes chiral-dependent electronic coupling at atomically structured gold surfaces as the basis for discrimination. However no structural characterization of electrode chirality is presented, no theoretical modeling is provided, and the SiO₂ coating description would seem to negate all this reasoning. I recommend clarifying all of this, and tempering the mechanistic claims and framing them as hypotheses unless supported by direct evidence.

<Reply>

We thank the reviewer for this important and insightful comment. We agree that our original discussion did not adequately distinguish between the timescale of self-assembled monolayer formation and the much shorter timescale of transient Au–thiol interactions. We also agree that the description of the SiO₂-coated nanogap architecture was insufficiently clear and that the mechanistic interpretation was previously stated too strongly.

A) Electrode architecture and the SiO₂ layer

The device contains a SiO₂ layer deposited over the electrode structure to reduce electrical noise and define the nanochannel sensing geometry. The nanogap is formed in situ by mechanically breaking a Au nanowire after device fabrication. Gold nanowires coated with SiO₂ are fractured at a constricted region of the nanowire, resulting in the exposure of the Au surfaces at the fracture surfaces. In practice, after the formation of a single-atom gold contact, the contact is ruptured to fabricate the gold nanogap electrodes. However, because the atomic-scale structure of the fractured gold surfaces cannot be directly characterized in detail, the detailed interfacial

structure remains unclear. The manuscript has been revised to clarify this point in the main text and the Experimental Section.

B) Potential contribution of transient Au–S coordination to cysteine dwell times

We agree that transient Au–thiol coordination can occur on timescales much shorter than those required for self-assembled monolayer formation. Therefore, the absence of monolayer formation cannot be used to exclude a contribution from transient Au–S interactions during individual tunneling events.

The present dataset was not designed to isolate or quantify transient Au–S coordination kinetics. Although cysteine dwell-time distributions are included in the amino-acid dataset, we do not have direct evidence allowing us to assign any particular component of the distribution specifically to Au–S coordination. We have therefore removed the previous argument based on monolayer-formation timescales and now state that transient coordination, adsorption geometry, and local interfacial interactions may contribute to cysteine dwell-time distributions and classification behavior.

C) Interpretation of the chiral discrimination mechanism

We agree that the current data do not establish chiral-dependent electronic coupling at atomically structured Au surfaces as a confirmed mechanism of discrimination. The electrode-tip structure was not directly characterized, and no theoretical modeling of the molecule–junction interaction was performed. In the revised manuscript, we therefore treat this mechanism as a hypothesis.

Specifically, we now state that the observed differences in tunneling-current waveforms are consistent with variations in transient molecule–electrode interactions during transport through the nanogap. Possible contributing factors include the local atomic structure of the mechanically generated junction, molecular orientation, steric effects, transient adsorption or coordination, and electrostatic interactions. Direct structural characterization of the junction apex, systematic control experiments with well-defined surface chemistries, and theoretical modeling will be required to determine the relative contributions of these factors and to establish the physical basis of chiral discrimination in this system.

[Revised Results, “Single-molecule detection of L- and D-form amino acids,” pp. 8–9; revised Discussion, pp. 19–21; revised Methods, “Fabrication of a device for single-molecule detection” and “Single-molecule measurements,” pp. 22–24; revised Fig. 1 legend.]

Results (pp. 8–9)

The SiO₂ layer was introduced to reduce electrical noise and define the nanochannel geometry. The nanogap was generated in situ by mechanically breaking the Au nanowire after device fabrication. As a result, gold is exposed at the fracture interfaces. However, the local atomic structure and chemical state of the mechanically generated junction apex, including the extent of SiO₂ coverage in the immediate sensing region, were not directly characterized in this study.

This interpretation suggests that the local nanogap environment may contribute to chiral sensing. However, the structural and chemical properties of the junction apex were not directly characterized, and the contributions of local surface structure, molecular orientation, transient adsorption or coordination, and electrostatic interactions remain to be determined.

Methods (pp. 24)

The SiO₂ layer was deposited onto the fabricated device structure to reduce electrical noise and define the nanochannel geometry. The nanogap junction was subsequently formed in situ by mechanical rupture of the Au nanowire. The local atomic structure and chemical composition of the resulting junction apex, including the extent of SiO₂ coverage at the immediate sensing region, were not directly characterized in this study.

Fig. 1 caption

When an amino-acid molecule passes through the nanogap between two electrodes, it can transiently modulate the tunneling current, generating a measurable single-molecule electrical signal. The L- and D-forms can yield distinguishable current–time waveforms, potentially reflecting differences in transient molecule–junction interactions, molecular orientation, and local interfacial effects within the nanogap.

Reviewer Comment: Issue 5. Concentration Regime and Practical Sensitivity for a Real Mission

The manuscript reports an LOD of 0.03 nM based on calibration against a blank signal frequency of $2.1 \pm 0.27 \text{ min}^{-1}$. However, after a careful analysis of their end-to-end analytical process for real samples like the ones you could encounter on a spaceflight mission, this value appears to represent an idealized instrumental detection threshold rather than a practically achievable identification limit under realistic conditions.

Several factors constrain the effective sensitivity:

1) Background-limited detection. A non-zero blank event rate establishes a statistical floor. At subnanomolar concentrations, analyte signal frequencies would approach background levels, and because the classifier lacks an “unknown” category, background events will necessarily be

assigned to amino acid classes. This inflates apparent signal and limits reliable identification near the stated LOD. Under these conditions, the effective detection limit is determined by the rate at which analyte-derived events can be distinguished from background events over the stable operational lifetime of the nanogap device.

2) Inability to extend acquisition time. The authors state that prolonged measurement does not substantially increase usable events due to adsorption and nanogap instability. This prevents improving sensitivity via longer integration, fundamentally constraining the operational detection limit.

3) Matrix and validation gap. Natural samples were analyzed at micromolar concentrations which are several orders of magnitude above the claimed LOD. No data are presented demonstrating accurate identification or D/L quantification at nanomolar levels in heterogeneous matrices. Therefore, the practical detection limit for complex samples has not yet been demonstrated and cannot be inferred from the calibration-based LOD reported here.

For these reasons, the quoted LOD of 0.03 nM is not directly relevant to mission-level trace detection as currently demonstrated and should be reframed accordingly. The authors should clearly distinguish between calibration-based detection limits under controlled conditions and the practical identification limits achievable in complex environmental samples.

<Reply>

We thank the reviewer for this important critique. We agree that the previously reported LOD of 0.03 nM represents a calibration-based instrumental metric obtained under controlled single-analyte conditions and should not be interpreted as a demonstrated practical identification limit for complex environmental samples or mission-relevant matrices.

We have revised the manuscript to clearly distinguish between the calibration-based detection performance of the nanogap measurement system and the practical performance required for molecular identification in heterogeneous samples.

First, a calibration-based LOD of 0.03 nM was obtained from the correlation between tryptophan signal frequency and concentration in a controlled aqueous solution. This value describes the concentration at which the signal frequency becomes distinguishable from the blank response under the specific calibration conditions used in this study. `

It does not imply reliable molecular identification, chiral classification, or D/L-ratio quantification at this concentration.

Second, the revised environmental-sample analysis incorporates a confidence-based rejection criterion. For each feature-extracted event, the maximum predicted probability across the predefined 11-species target panel was calculated. Events with a maximum probability < 0.5 were classified as unknown/rejected and excluded from the targeted compositional and D/L-ratio analyses. Thus, the revised workflow no longer assigns all detected events to one of the predefined amino-acid classes. Nevertheless, the process blank still retained a small fraction of high-confidence target assignments, and these retained blank assignments showed class-dependent enrichment. These findings demonstrate that background events and non-target matrix components can still affect practical molecular identification, particularly in low-abundance samples.

Third, the environmental extracts analyzed in this study were reconstituted at estimated amino-acid concentrations of 1–100 μ M for nanogap measurements. These concentrations are 4–7 orders of magnitude higher than the calibration-based LOD. Therefore, the present environmental-sample experiments do not demonstrate reliable amino-acid identification or D/L-ratio quantification at nanomolar concentrations in chemically complex matrices.

We have also clarified that the measurement time used for the calibration experiments was 15 min per concentration, and that the present study did not evaluate how longer acquisition times would affect practical identification performance in complex samples. Although increasing the number of detected events may improve statistical precision under certain conditions, the effective operational time of a nanogap measurement can be constrained by junction stability, surface effects, and sample-dependent changes in event frequency. These operational factors require systematic evaluation in future studies.

Accordingly, we have revised the manuscript to state that the 0.03 nM value is a calibration-based instrumental LOD, whereas the practical limits for molecular identification and quantitative chiral analysis in complex environmental samples remain to be established through matrix-matched calibration, contamination-controlled validation, increased event statistics, and testing under mission-relevant operational conditions.

[Revised Abstract, p. 3; revised Introduction, p. 5; revised Results, “Single-molecule detection of L- and D-form amino acids,” pp. 8–9; revised Results, “Application of chiral classification to amino acids extracted from samples of the Murchison meteorite and the Atacama Desert,” p. 20; revised Methods, “Single-molecule measurements,” p. 25; revised Methods, “Amino-acid extraction from astrobiologically relevant samples,” p. 28; revised Discussion, pp. 21–22.]

Abstract, p. 3;

Life on Earth uses L-amino acids and D-sugars, whereas abiotic processes produce equal abundances of both enantiomers. Detecting enantiomeric excess among molecules of biological interest has therefore been proposed as a potential biosignature for life beyond Earth. **However,**

robust electrical methods for determining the L/D ratios of amino acids and sugars at low concentrations, particularly in chemically complex samples, remain limited. Here we present a single-molecule approach based on quantum tunneling currents that discriminates between L- and D-form amino acids and quantifies their ratios without using relying on optical detection or recognition molecules. Individual L- and D-forms were distinguished with over 80% accuracy, and in a blind test of 39 amino acids, identities and chiralities were identified with >50% accuracy, far above random chance. Mixtures of four amino acids with variable chiralities were quantified with >66% accuracy. Finally, Murchison meteorite and Atacama Desert extracts were analyzed using a targeted 11-species amino-acid panel. Confidence-based filtering distinguished high-confidence target assignments from low-confidence events, revealing sample-dependent contributions from background and non-target signals. The resulting high-confidence compositions captured selected major compositional features observed via LC–MS while highlighting limitations in low-abundance and matrix-rich samples. These results establish a targeted single-molecule electrical approach for chiral amino-acid analysis and define important requirements for its application to complex astrobiological samples.

Methods, “Amino-acid extraction from astrobiologically relevant samples” (p. 28).

Amino acids extracted from astrobiologically relevant samples.

Amino acids were extracted from the Murchison meteorite and from two sites in the Atacama Desert (North of Antofagasta (NOA) and Diego de Almagro (DGA)). For each extraction, ~1.2 g of dried sample was mechanically homogenized and subjected to hot-water extraction at 100°C for 24 h, followed by centrifugation, multiple washing steps, and drying. The NOA and DGA extractions were performed in triplicate. A portion (45%) of each extract underwent vapor-phase acid hydrolysis at 150°C for 3 h in 6 M HCl. Amino acids were further purified via cation-exchange chromatography, and the eluates were dried and stored before analysis. The resulting acid-hydrolyzed samples were reconstituted in Milli-Q ultrapure water and analyzed via both liquid chromatography–mass spectrometry (LC–MS) and single-molecule nanogap detection (Supporting Information S5, Fig. S13). For single-molecule analysis, dried samples were rehydrated in Milli-Q ultrapure water to yield amino-acid concentrations of 1–100 μ M, which are optimal for nanogap measurements. Owing to the higher total amino-acid content, Murchison extracts were diluted to a final volume of 10 mL, whereas NOA and DGA extracts were rehydrated in 1 mL. All single-molecule measurements were conducted using freshly prepared 50- μ L aliquots to ensure experimental consistency. Further details on the samples, amino-acid extraction workflow, and LC–MS and single-molecule analytical procedures are provided in Supplementary Information Sections S3–S10, Figures S13–S18, and Table S2.

Discussion (pp 20–21)

Herein, we evaluated nanogap-based single-molecule analysis using amino-acid extracts from astrobiologically relevant samples. Confidence-based filtering enabled separation of high-confidence target assignments from low-confidence events within a predefined 11-species amino-acid panel. The process blank showed a low fraction of high-confidence target assignments, whereas the Murchison sample exhibited a substantially higher retained target fraction. The DGA and NOA samples showed intermediate target fractions, indicating that process-derived background signals and matrix-dependent non-target events contribute more strongly to the results obtained for some environmental extracts.

The confidence-filtered compositional and D/L ratio analyses captured selected major features observed via LC–MS, particularly for the Murchison, and showed analyte- and sample-dependent discrepancies. The present method should therefore be interpreted as a targeted, high-confidence identification approach rather than as comprehensive molecular profiling or absolute quantification of amino-acid mixtures in environmental samples. Future work should incorporate representative blank-derived and nontarget molecular datasets, matrix-matched calibration experiments, and open-set recognition approaches to improve discrimination between target analytes, background signals, and chemically unrepresented events.

In summary, we have revised the manuscript to present the 0.03 nM calibration-based LOD as an idealized benchmark under controlled single-analyte conditions, and we explicitly state that the practical identification limit for complex environmental samples has not yet been fully established. We have added a new paragraph to the Discussion outlining the remaining steps—particularly validation at nanomolar concentrations in spiked complex matrices—required before this technology can be confidently applied in planetary exploration contexts. We believe that this revised framing provides a more accurate and scientifically rigorous representation of the current state and future prospects of the method.
