nature portfolio

Peer Review File

Sequential reading of a stepwise shortened peptide immobilized on nanopore

Corresponding Author: Professor Shuo Huang

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

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

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
This is a very compelling manuscript reporting the development of transient pore analyte looping (tPAL), a nanopore-based approach towards peptide sequencing. By dual modification of an MspA nanopore with a Ni-NTA and a peptide, the authors enable acquisition of electrical signals from multiple Ni-peptide interactions. Sequential enzymatic cleavage using an aminopeptidase enables progression through the sequence, generating a signal with sequence-dependent current 'steps', enabling discrimination of amino acid differences, post-translational modifications and unnatural amino acids.

In this work, the authors functionalise a single protomer of MspA with two different moieties, to facilitate functionalisation of MspA with a peptide for sensing and a Nickel ion to act as the sensor. A Ni-NTA and a reactive handle to facilitate peptide immobilisation, this is done by introducing the unnatural amino acid, AzK, to facilitate peptide immobilization through SPAAC and modifying a cysteine with maleimide-C3-NTA. To my knowledge this work is the first published example where a nanopore has undergone site-specific functionalisation with two orthogonal groups. Other methods have been developed for resampling of peptides to improve the confidence of inferences made.

Brinkerhoff, H., Kang, A. S. W., Liu, J., Aksimentiev, A., & Dekker, C. (2021). Multiple rereads of single proteins at single–amino acid resolution using nanopores. Science, eabl4381. https://doi.org/10.1126/science.abl4381

Motone, K., Kontogiorgos-Heintz, D., Wee, J., Kurihara, K., Yang, S., Roote, G., Fox, O. E., Fang, Y., Queen, M., Tolhurst, M., Cardozo, N., Jain, M., & Nivala, J. (2024). Multi-pass, single-molecule nanopore reading of long protein strands. Nature, 1–8. https://doi.org/10.1038/s41586-024-07935-7

The tPAL method presents an alternative solution for peptide resampling. It would be useful for comparisons to be made with these existing methods to allow readers to better understand the advantages/disadvantages of the various resampling methods.

Fig 1c – description in figure legend and accompanying text (98-103) requires some additional information. It would be useful to clarify what method was used to separate the different protein species – SDS-PAGE/BN-PAGE etc. or include additional relevant information in the methods.
The manuscript is clearly written, well-structured and has a logical flow. Figures are informative and technical terms are well defined. The methods are generally well described in the manuscript and accompanying supplementary information. The electrophysiology data included (Figs 2-6) are well presented, supplementary information shows reproducibility across multiple nanopores.
Previous publications on the topic of single-molecule peptide sensing using nanopores have substituted all 20 proteinogenic residues into a model analyte to compare the discriminative power of their respective sensing methods.

Motone, K., Kontogiorgos-Heintz, D., Wee, J., Kurihara, K., Yang, S., Roote, G., Fox, O. E., Fang, Y., Queen, M., Tolhurst, M., Cardozo, N., Jain, M., & Nivala, J. (2024). Multi-pass, single-molecule nanopore reading of long protein strands. Nature, 1–8. https://doi.org/10.1038/s41586-024-07935-7

Ouldali, H., Sarthak, K., Ensslen, T., Piguet, F., Manivet, P., Pelta, J., Behrends, J. C., Aksimentiev, A., & Oukhaled, A. (2020). Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. Nature Biotechnology, 38(2), 176–181. https://doi.org/10.1038/s41587-019-0345-2)

It would strengthen the manuscript to extend the scope of the work presented in either Figure 4 or supplementary Figure 11, including further substitutions in order to demonstrate discrimination of all 20 (or 19 without cys) proteinogenic amino acids and show that they are all separable in this sensing regime.

Features of events were shown to be reproducible across replicates (supplementary tables 5 & 10). Error bars are included in all graphs where appropriate. The authors report a Cubic SVM model (supplementary table 6 & supplementary Figure 18) able to discriminate a peptide and peptides representing stepwise cleavage products to high accuracy (99.5%). Both test accuracies and those from cross-validation are reported, along with the results from a variety of other models. In the legend for supplementary table 6 it is stated that datasets of 400 events were generated for each peptide; however, it is not clear how many nanopores this came from. The peptides discriminated by the Cubic SVM model vary in both length and N-terminal amino acid, thus I think some clarification is needed for the following wording: “illustrating the tPAL strategy’s efficacy in distinguishing single amino acid differences within peptide sequences” (240-242).

It would be useful to clarify the statement “Additionally, for de novo peptide sequencing, a reference database and corresponding AI-driven algorithms are essential for decode nanopore events into the corresponding peptide sequences.” (414-416) The use of a reference database isn’t sequencing de novo.

As the tPAL approach is currently limited to one peptide per experiment (of a given length), it would also be useful for the authors to comment on how this limitation could be overcome both with respect to generating training data for machine learning but also for enabling the method to address biological questions.

Overall, this manuscript makes a major conceptual and technical contribution to the field of nanopore protein sequencing by introducing a new method for peptide analysis. I strongly support publication after minor clarifications and extensions as noted above.

Referee #2

(Remarks to the Author)
Chen and colleagues describe a method for peptide identification using nanopores. The system relies on a heterooctameric MspA nanopore in which one protomer carries two modification sites: the first attaches a Ni-NTA moiety, and the second attaches a maleimide moiety. The latter enables the peptide to be tethered to the nanopore, while the former allows recognizing to the N-terminus of the attached peptide. Each tested peptide yields a characteristic current blockade, which is likely to be strongly influenced by the identity of the terminal amino acid. An aminopeptidase anchored to the membrane then cleaves the last amino acid, providing a new peptide’s recognition by the nanopore. This process is repeated until all amino acids are removed.

This approach is an ingenious extension of prior work from the Huang group. The manuscript is well written, and the data are convincing. However, I have reservations about the potential of this method to deliver a practical sequencing technology. This is for several reasons:

1. Attachment to the nanopore is required for each peptide. Currently, this is accomplished using cysteine residues. This is problematic because relatively few peptides contain cysteine, and if the cysteine is not at the C-terminus, the reading is incomplete. The authors claim that nanopores can be engineered to react with other amino acids, but such chemistries exist for only a limited subset of residues. Moreover, signals from residues extending from the attachment point toward the C-terminus would still be unobservable. The only plausible solution I can envision is to develop a chemistry that enables selective reaction at the peptide’s C-terminus with the nanopore; however, this remains an unresolved chemical challenge.

2. Throughput is limited because only one peptide can be analyzed per nanopore at a time. This would necessitate forming and breaking lipid bilayers after each measurement, leading to very low throughput.

3. The method does not provide straightforward peptide sequencing. Prior work suggests that more than ten amino acids can influence the signal, so it is highly likely that the observed blockade results from multiple residues occupying the recognition site, rather than solely the terminal amino acid. While the authors show that terminal histidine and proline can generate distinctive signals, other amino acids yield current blockades that do not appear specific to the last residue. The noise of the blockade may carry information about the entire peptide, and the kinetics of amino acid cleavage could correlate with the last few residues. Thus, the overall ionic signature as the peptide transitions from full length to fully digested may be unique to a given peptide, but this has not been demonstrated. Nonetheless, given the existence of millions of peptides in the human proteome, the required training set to address all possible sequences is likely to be enormous. To begin addressing this challenge, the authors should undertake systematic testing of peptides. One possible approach would be to examine sets that share the first two residues (e.g., LR) but differ at the third position (i.e., 20 LRX peptides). If the current delta is the same

across all 20 peptides, the signal might reflect the L→R transition; subsequent experiments could then explore other residue combinations.

An additional point that could be explored with more experiments is the effect of peptide length. The authors tested only three peptides and did not address whether longer peptides—with potential structural features—would influence the signal, nor whether the peptidase can efficiently access longer substrates.

In summary, the results presented appear too preliminary to convincingly propose a method for sequencing peptides with nanopores.

Referee #3

(Remarks to the Author)
Summary
Protein sequencing has recently emerged as a central research focus within the nanopore community. Existing approaches can be broadly categorized into protease-assisted sequencing, fingerprinting, and strand sequencing, with strand sequencing generally considered the most promising for broad applications. Protease-assisted and fingerprinting methods, while valuable, often fall short of true sequencing due to capture preferences and limited generality across proteomes. In this context, Chen et al. present a strategy they term tPAL (transient Pore Analyte Looping). The authors engineer a hetero-octameric MspA nanopore with two bio-orthogonal reactive sites: one for NTA–Ni modification to capture the peptide N-terminus, and another for covalent tethering of the peptide near the pore constriction. Coupled with a cholesterol-modified aminopeptidase, this enables stepwise N-terminal trimming, multiple rereads of the same analyte, and discrimination of single amino acids, selected post-translational modifications (PTMs), and unnatural residues. The manuscript is clearly written. The technical achievement of constructing a dual-functionalized pore and demonstrating in-pore enzymatic stepping with single-residue resolution is commendable and will be of interest to the nanopore community. However, while these advances represent a step forward, I have reservations about the conceptual innovation, mechanistic clarity, and general applicability, which need to be addressed before the work can reach the bar for Nature. I recommend major revision.

Major Concerns
1. Enzyme dependence and generality. The method relies almost exclusively on sgAP, which shows sequence preferences and stalls at certain motifs (e.g., Pro or phosphorylated residues). This constrains its utility for proteome-wide sequencing. Demonstrating activity with an enzyme panel or cocktails would substantially strengthen the generality of the approach.
2. Error rates and blinded decoding. Reported high accuracies are derived from curated datasets. No blinded decoding of mixed peptides, no confusion matrices, and no tests under varied experimental conditions are presented. These experiments are essential to evaluate real-world sequencing performance.
3. PTM coverage. The study demonstrates detection of a small number of PTMs, but does not assess isobaric or chemically similar modifications (e.g., Lys acetylation vs. methylation). Expanded PTM analysis is needed to support broader claims.
4. Sensing mechanism. Several aspects of the physical basis for current signatures remain unclear:
a) Which forces (electro-osmotic flow, electrophoretic force, or both) govern loop configuration inside the pore?
b) How do signals compare between tethered and untethered peptides, given their different conformational freedoms?
c) Although the authors note multiple residues may influence the current, Figure 5 suggests signals may often be dominated by the N-terminal residue. Clarification of how many residues collectively define a signal is crucial.
d) Can tethered peptides enter the pore without NTA–Ni^{2+} coordination, and if so, how often? The observed transient spikes need to be explained in this context.
e) If PEG–peptide conjugates can reach the sensing region without coordination, why do kon and Kd values differ across amino acids? For example, cysteine has an unusually high kon.
f) The reported decrease in tPAL frequency with peptide or linker length raises the question of whether spike events become more frequent under these conditions.
g) Does peptide length affect the magnitude or fluctuations of the signal for a given residue? If so, this introduces a length-dependent complication; if not, the role of loop tension and conformation should be clarified.
h) Claims that tPAL events resemble stochastic sensing (Fig. S13) should be tested more broadly, and the contribution of loop tension explicitly considered.
i) Signals from PEG24- vs. PEG4-linked peptides are similar (Fig. S14), raising concerns about whether linker length meaningfully affects conformation. In addition, Fig. S15a shows unexpectedly flat signals after binding, which requires explanation.
j) Differences in koff values (Fig. S21) suggest more complex peptide configurations than described.
k) Further experiments with residues such as histidine at positions beyond the N-terminus (e.g., position 2 or 3; Fig. S30) would clarify how many residues are simultaneously detected.

Minor Points
5. To illustrate changes in ionic conduction before and after chemical modification, all I–V curves should be plotted together for direct comparison.
6. The peptide modification process involves two steps: NTA–Ni^{2+} coordination at the N-terminus and cysteine–maleimide conjugation at the pore wall. Kinetic information on both steps, including the rate-limiting factor, should be provided to assess feasibility.
7. While multiple rereads of a peptide are emphasized as a strength, they may also obscure resolution when residues are

identical or generate similar signals. This limitation should be acknowledged.
8. The apparent enhancement of sgAP activity arises from immobilization on the lipid membrane, which increases local concentration, rather than intrinsic activity changes. This should be clarified.
9. In line 353, the correct reference is Fig. 5d and h, not Fig. 6d and h.

Referee #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
The authors have responded comprehensively to my initial comments and have substantially strengthened the manuscript through the addition of new experimental data and improved technical clarification.
The additional SDS–PAGE methodological details are appreciated and improve reproducibility. However, clarification is needed regarding the electrophoresis conditions. The manuscript states that a 10% SDS–PAGE gel was run at 160 V for 16 h, which seems unusually long at that voltage and may represent a typographical error. Please clarify the actual running conditions used. In addition, please specify the SDS–PAGE buffer system employed (e.g. Tris–glycine–SDS or Tris–tricine)
The effort placed into performing the suggested experiments to assess discriminative power of the system is appreciated. The revised manuscript now convincingly demonstrates discrimination of all 20 proteinogenic amino acids and utility of multi-voltage rereading to enhance resolution and classification accuracy.
My concerns regarding dataset construction and statistical independence have been addressed by clarifying that machine-learning datasets were assembled from multiple independent pores.
The revised terminology surrounding “de novo” sequencing is also more precise, and the scope of the decoding demonstration is now appropriately framed.
Limitations related to throughput and scalability are acknowledged by the authors and are reasonably positioned as future engineering challenges.
Overall, the revisions largely address my prior concerns. Pending further clarification of points raised above, I would be supportive of publication.

(Remarks on code availability)

Referee #2

(Remarks to the Author)
Chen and coworkers have addressed all the questions I have raised. It is commendable to see how well the authors argued their points and how much extra work has been done to improve the manuscript. There is no doubt this work ranks highly among nanopore papers produced recently.

My concern, however, remains. as they are exclusively on the impact of this work on protein sequeincing. I think that journals of the calibre of Nature represent an important vehicle for science dissemination, not only reflecting the technical ability of the authors, but also the potential impact of the technology that is presented.

Therefore, I focus my comments exclusively on the innovative step, feasibility of this technology, on the challenges to make this technology a reality.

Innovation.
It has long been demonstrated that nanopores can distinguish between molecules that have a similar size as the nanopore constriction. The authors have previously shown that MspA (the nanopore used here) can distinguish between all 20 amino acids and PTMs. Hence, it is not surprising that this approach can also do the same.

The novelty lies in attaching a peptide to the nanopore. This could allow single-molecule recognition of the terminal amino acid (see comment under’recognition’), but it brings many other issues, as underlinded below.

Throughput.
This seems to me the major challenge. Peptides need to be attached to the nanopore and read once. Reversible chemistry with pre-inserted nanopores is required to obtain high thoughput, but it needs to be developed (and not using a disulfide bond strategy). It will alsol require buffer exchange to operate which will complicated the operations and increas operational time, and must be fast. If reversible chemistry cannot be developed, nanopores need to be reacted with the target peptides and re-reconstituted into membranes after every run. This would require many millions of nanopores simultaneously that record data with high resolution. Furthermore, it has to be demosrated that the nanopore with an attached peptide can insert

(especially if the peptide is peptide, will allow insertion).

C-terminal tagging.
This is yet unsolved. The authors have rightfully indicated a few examples where C-terminal tagging has been used. However, that method is by no means generic or specific for all peptides. Lys-C tagging can also be used, but it would probably result in 50% of the peptides to be unusable, unless a lysine-specific tagging is identified. Nonetheless, I agree with the authors that this challenge might be eventually solved by the community at large.

Recognition.
Several previous works have shown that 20 amino acids and their modifications can be identified by nanopore currents. The main authors have published recently the discrimination of 20+ amino acids using a very similar nanopore as indicated here.

The main issue is that the recognition is not of the last amino acid but of several amino acids at the constriction. The authors mention only three amino acids are recognised, but this is unlikely. Previous work with both DNA and proteins suggested several more amino acids are likely recognised, and the number of recognised amino acids will depend on their sequence.

But the main issue is the k-mer in combination with the short reads. Even if we assume that the peptidase will work in favour of sequencing (see next point), the combination of possible sequences is very large. For intact proteins passing through a nanopore, this issue is somehow mitigated by the length of the protein. Each amino acid is read multiple times (like for DNA sequencing) and the challenge is that of having enough information to identify one protein over a few thousand. Here, the number of possible combinations appears simply too large (multiples of millions). The authors have attempted building a simple classifier in the rebuttal, but it did not address the point.

Amino peptidase reading.
This approach requires that the 1) terminal peptide is cut one amino acid at a time, 2) the amino peptidase disengages from the peptide, 3) the terminal amino acid is read once, 4) the amino peptidase binds again to the terminus of the peptide and cuts it once more.
Previous work indicated that amino peptidases do not cut-disingage-cut-disingage… but they bind and then cut several peptides at the same time. However, let us assume an amino peptidase can be engineered to cut once and then disingage the peptide. In this case, there would still be a competition between nanopore readout and re-engaging with the peptidase. This could be minimised by having a very low peptidase activity, effectively reading the peptide many times. However, this would prevent reading two identical consecutive amino acids. I cannot see how to identify consecutive amino acids that are identical. The authors should try to design an experiment. Can peptides with stretches of, say, DD, DDD, and DDDD be identified? Maybe if the k-mer size is actually large, this difference can be identified.

Peptide length.
The authors mentioned they used a peptide of 22 amino acids, but I could not find the data (Figure S28 shows the sequential cleavage of LRGESC, maybe they indicate S19? There there is a 22 amino acid peptide, but only three steps are observed. The authors mention that denaturing conditions can be used, but this would inactivate the enzyme. The structure of long peptides will affect the k-mer recognition. The authors should clearly state if they observe 22 cleaveage steps of Pep22.

In conclusion, despite the best effort from the authors to improve and explain their method, I maintain the view that it appears extremely unlikely this approach will allow peptide sequencing.
The main issues I believe the authors should address are:
1. It has not been demosntrated that long peptides can be addressed by this approach -i.e. the authors should show that the 22mer peptide can originate 22 amino acid steps (not necessarialy sequenced)
2. Homopolymeric sequence should not be possible if the k-mer is small (an experiment can be designed for studing this)
It should be noted that if the k-mer reads more than 3 amino acids, homopolymeric peptides should become less of an issue. However, peptide sequencing would become very challenging given the astronomical possible combinations of peptides. However, it would not be impossible, as in theory the sequential cleveage by the enzyme would allow to read individual amino acids multiple time.

(Remarks on code availability)

Referee #3

(Remarks to the Author)
I have reviewed the revised version and the associated responses. Relative to the prior version, the manuscript has been substantially expanded in experimental scope and analysis. The authors present (i) a dual-functionalized hetero-octameric MspA pore enabling peptide immobilization and repeated N-terminal interrogation via NTA–Ni coordination (“tPAL”), (ii) stepwise enzymatic cleavage to generate sequential, stage-dependent signal changes, and (iii) broader demonstrations including PTMs (notably lysine acetylation vs trimethylation), phosphorylated substrates (with an enzymatic workaround), unnatural amino acids, D-peptides (sensing only), repeating-sequence peptides, and a machine-learning pipeline for stage classification and proof-of-concept sequence reconstruction. Overall, the revised work is data-rich and more convincing technically than the original submission. I recommend acceptance.

(Remarks on code availability)

Referee #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

(Remarks on code availability)

Version 2:

Reviewer comments:

Referee #2

(Remarks to the Author)
The authors have demonstrated excellent argumentative skills, thoroughly addressing each comment. As previously mentioned in my review, however, my evaluation goes beyond the technical assessment of this work or the argumentative skills of the authors, but rather focuses on the novelty of their approach and its applicability to protein sequencing.
As previously noted, numerous nananopore experiments have demonstrated the ability to distinguish between individual amino acids. Furthermore, unnatural amino acids have been introduced into nanopores, and the MspA-NTA-Ni nanopore has been utilized in numerous articles authored by the main author.
Therefore, the innovative step involves attaching a peptide to the nanopore and reading the initial amino acid through current recording.
Nevertheless, I believe this approach presents several practical challenges:
1. The peptide must be attached to the nanopore, and the nanopore-peptide must be inserted into a membrane. While the authors have demonstrated that small peptides allow insertion, it is uncertain whether large peptides will be tolerated. This has not been substantiated by additional experiments.
2. The nanopore signal is influenced by multiple amino acids, rendering sequencing a complex task. In their initial version, the authors implied that the signal is dominated by the last amino acid, implying the possibility of sequencing. Following my inquiry regarding homopolymeric peptides, the authors acknowledged that the signal is influenced by several amino acids, indicating that sequencing will be highly challenging. This issue is particularly relevant for small peptides, as natural peptides possess a vast sequence space. A recent study from BGI-Shenzhen, which employed a different but sequencing-compatible system (10.1038/s41467-026-69628-1), effectively illustrates this problem.
3. Long peptides exhibit a degree of structural organization, rendering the N-terminus less accessible. In my previous review, I requested the authors to provide data on the sequencing of a long peptide. Regrettably, they opted for an artificially constructed sequence: EQTREQTREQTREQTREQTREQTRC. This is disappointing, as the peptide lacks folding and comprises only four amino acids.
Most of these issues I have raised could have been resolved if the authors had characterized a long peptide with a random sequence. Notably, I previously requested in my review that the authors address the issue of secondary structure in long peptides. However, they chose to test EQTREQTREQTREQTREQTREQTRC, a non-natural peptide with a simplified amino acid composition and no secondary structure.
Consequently, despite their arguments, the authors have neglected to address the primary concern of this work. This is disappointing, as the manuscript currently presents a misleading portrayal, concealing the main issues with this approach.
To address this, I propose an experiment. The authors should attach a peptide of 20 or more amino acids derived from a lys-C digest of a natural protein (the author can select the peptide). The peptide can include an additional cysteine at the C-terminus to facilitate nanopore conjugation.
Then, they should demonstrate that a nanopore with such a peptide attached can insert into a membrane and that the entirety of the peptide can be read by the nanopore, distinguishing the twenty or more steps.
It is important to note that this is a simplified experiment that does not yet demonstrate sequencing. The C-terminal attachment is not addressed (the authors can use a cysteine at the C-terminus), and I am not requesting that the peptide be sequenced, but rather that each step can be distinguished.
If the authors can successfully perform this experiment, I would recommend that they submit the manuscript for publication.

(Remarks on code availability)

Version 3:

Reviewer comments:

Referee #2

(Remarks to the Author)
The authors have conducted the experiment I requested, selecting the 22-mer peptide ANWEAQQRILEQQNSSRTLEKC from human septin-7. The peptide contains an aromatic residue (tryptophan at position 3) alongside isoleucine and two leucines, which support a modest hydrophobic core and some degree of helical structure—consistent with its predicted α-helical conformation in the intact septin-7 protein. To more rigorously address the folded issue, a peptide enriched in

multiple aromatic (Trp, Phe, Tyr) and aliphatic hydrophobic residues (Leu, Ile, Val, Pro) would have provided a more stringent test of the system's ability to handle structured substrates, where N-terminal accessibility to the nanopore and the aminopeptidase may be genuinely compromised. Nevertheless, in light of the data presented, I think this peptide provides a reasonable demonstration that a peptide with a scrambled sequence can be addressed by the system.

The authors do not, however, provide a detailed interpretation of the ionic current traces in the rebuttal, which is frustrating. My interpretation by inspection of the traces in the figure provided reveals that the ionic current signatures for two chemically identical glutamine pairs differ significantly from one another. If true, this clearly indicates that the nanopore does not read a single terminal amino acid in isolation; rather the ionic current at each cleavage stage encodes the combined contribution of several adjacent residues —a k-mer effect. On one hand, this findings suggests the system is capable of reading short homopolymeric sequences. On the other hand, it reveals that the signal reflect several amino acids simultaneously, which complicates the system's ability to sequence generic peptides—a point I return to below.

An experiment with a scrambled amino acid sequence is therefore important and should be presented in the main text with figures, as it would provide a clear picture of the capabilities and current limitations of this methodology.

As it stands, in the section "Direct decoding of peptide sequence," the authors do acknowledge the existence of a k-mer effect:
1) "Based on the insight that peptide events are predominantly influenced by the first few amino acids at the N-terminus" (lines 348-349).
2) "…suggesting that the first three N-terminal amino acids play a dominant role in determining event features" (lines 344-345).
3) "Notably, the tPAL events in each stage encode information not only about the N-terminal amino acid of that stage but also about few downstream residues" (lines 361-362).

However, on a first reading, these statements are unclear, as no data are shown for generic peptides—only a cryptic and superficial Figure S54. The new result with the natural-sequence peptide—currently buried in the SI with minimal discussion —makes the meaning of these statements clear, and also explains why all sequencing demonstrations in this manuscript employed repeat-sequence peptides. The authors have been validating their system precisely under conditions where its most fundamental limitation is invisible. This is an important finding that the community deserves to understand fully, and it should not be concealed in a supplementary figure with minimal discussion.

Therefore, the authors should present in a main-text figure an experiment—which could be the new data recently collected—showing the signal from a scrambled peptide sequence, described in detail. They should explicitly state that the same amino acid is read differently depending on its sequence context. This is a key feature that the reader must grasp while reading the article. The authors should then openly discuss whether protein sequencing can be achieved with this approach, and how. This should not be buried in an eloquent but vague passage, nor relegated to a supplementary figure.

The issue with K-mer in peptides.
If, as superficially mentioned, only three amino acids contribute to each signal—which is already an optimistic estimate, given that the k-mer for DNA nanopore sequencing is typically 5 and estimates for peptide k-mers place it around 8—then $20^3 = 8{,}000$ distinct signal states must be deconvolved to sequence an arbitrary peptide from each single read. This makes sequencing highly challenging, given the vast combinatorial space of natural peptide sequences. If the effective k-mer is larger, sequencing may not be feasible at all. Nevertheless, as demonstrated in DNA nanopore sequencing, the step-wise nature of enzymatic cleavage simplifies the overall decoding problem. Furthermore, repeated reading of the same (cleavage) position—enabled here by multiple binding/unbinding events at each stage—should reduce stochastic error in k-mer calling.

(Remarks on code availability)

Response Letter

Reviewer #1:

Review Comments:

This is a very compelling manuscript reporting the development of transient pore analyte looping (tPAL), a nanopore-based approach towards peptide sequencing. By dual modification of an MspA nanopore with a Ni-NTA and a peptide, the authors enable acquisition of electrical signals from multiple Ni-peptide interactions. Sequential enzymatic cleavage using an aminopeptidase enables progression through the sequence, generating a signal with sequence-dependent current 'steps', enabling discrimination of amino acid differences, post-translational modifications and unnatural amino acids.

In this work, the authors functionalise a single protomer of MspA with two different moieties, to facilitate functionalisation of MspA with a peptide for sensing and a Nickel ion to act as the sensor. A Ni-NTA and a reactive handle to facilitate peptide immobilisation, this is done by introducing the unnatural amino acid, AzK, to facilitate peptide immobilization through SPAAC and modifying a cysteine with maleimide-C3-NTA. To my knowledge this work is the first published example where a nanopore has undergone site-specific functionalisation with two orthogonal groups.

Response:

We sincerely appreciate Reviewer 1 for the meticulous evaluation and constructive feedback on our manuscript. It is a great honor to receive your recognition of the significance of our work. As the reviewer accurately pointed out, one of the key innovations of our study lies in the dual functionalization of nanopore and the spontaneous interaction generated thereof, which, to the best of our knowledge, has not been reported previously, confirming the innovation of the tPAL technique. Notably, we dedicated significant effort to the development of this unique nanopore configuration with orthogonal reactive sites and to selecting the appropriate chemical strategy for successful dual modification. Your recognition of this innovation is a great encouragement to us.

We also sincerely acknowledge your valuable technical comments on our work. In this response, we have systematically addressed each of your comments. For your convenience, detailed responses to the comments are provided below in a point-by-point format.

Review Comments:

Other methods have been developed for resampling of peptides to improve the confidence of inferences made.

Brinkerhoff, H., Kang, A. S. W., Liu, J., Aksimentiev, A., & Dekker, C. (2021). Multiple rereads of single proteins at single-amino acid resolution using nanopores. Science, eabl4381. https://doi.org/10.1126/science.abl4381

Motone, K., Kontogiorgos-Heintz, D., Wee, J., Kurihara, K., Yang, S., Roote, G., Fox, O. E., Fang, Y., Queen, M., Tolhurst, M., Cardozo, N., Jain, M., & Nivala, J. (2024). Multi-pass, single-molecule nanopore reading of long protein strands. Nature, 1-8. https://doi.org/10.1038/s41586-024-07935-7

The tPAL method presents an alternative solution for peptide resampling. It would be useful for comparisons to be made with these existing methods to allow readers to better understand the advantages/disadvantages of the various resampling methods.

Response:

We sincerely appreciate Reviewer 1's constructive comment on resampling and your suggestion to include a technical comparison between tPAL and existing peptide resampling strategies. Following a thorough analysis of literatures listed by the reviewer, we have outlined the distinctions between prior approaches and our strategy as follows.

1. **Configuration:** tPAL achieves multiple resampling of each measurement step. However, the other two strategies resample the reading of the whole peptide for multiple times.

2. **Chemical vs. Enzymatical:** Although an aminopeptidase is used to sequentially cleave the N-terminal amino acids, this enzyme does not participate in the peptide resampling. Instead, the resampling is based entirely on well-designed chemical interactions, making the process more controllable and consistent. In contrast, other strategies rely on enzymatic activity for resampling, which is inherently stochastic. A major drawback of such methods is the frequent and random occurrence of peptide backsliding, which is undesirable for accurate sequence decoding (**Figure RL 1.1**). Additionally, resampling can easily be interrupted if enzyme binding fails stochastically. In contrast, with tPAL, peptide backsliding is technically impossible, and resampling can be carried out as many times as needed, easily exceeding thousands of repetitions. The stochastic nature of enzyme-driven approaches may also lead to the skipping of specific stages or inconsistent read lengths across cycles.

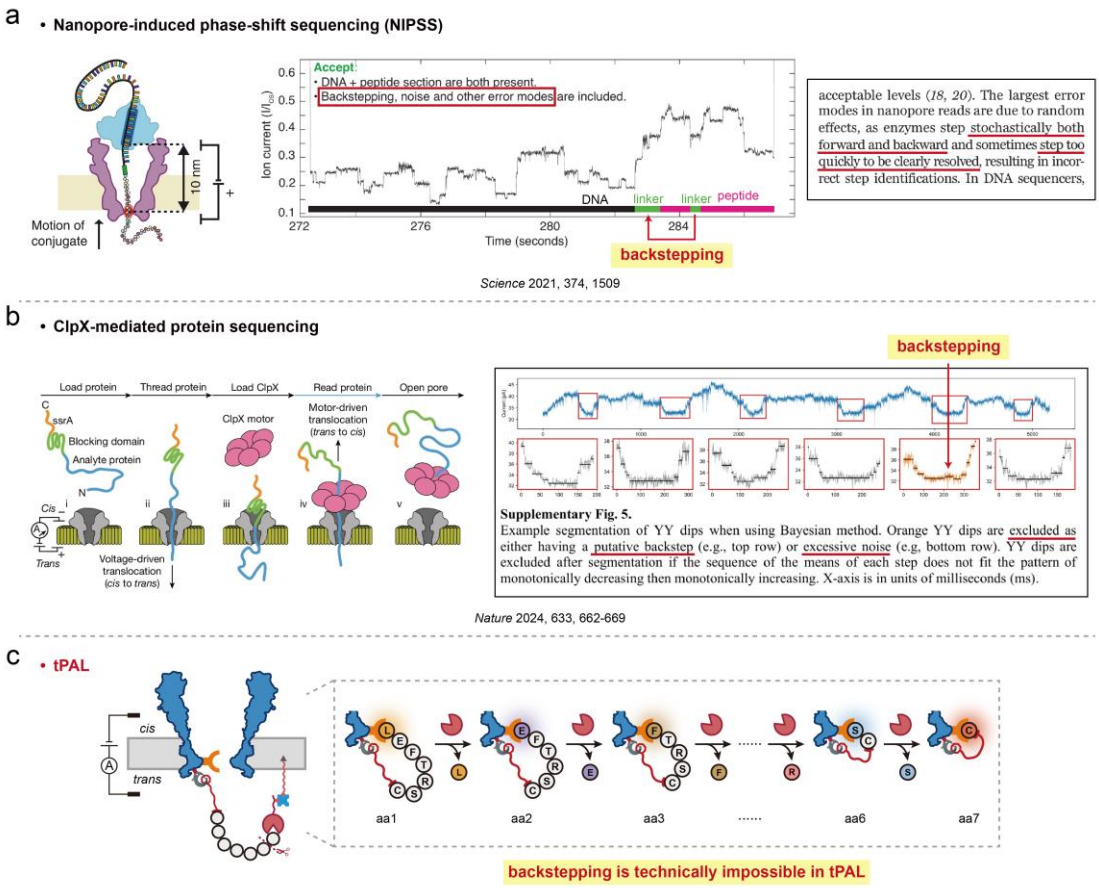


Figure RL 1.1. The technical advantage of tPAL method regarding unidirectional single amino acid stepping. Strand sequencing methods, such as NIPSS and ClpX-mediated protein sequencing, often face challenges related to strand backsliding. In contrast, the tPAL method demonstrates a notable advantage in achieving unidirectional single amino acid stepping through stepwise enzymatic digestion.

3. **With/Without chemical constraint:** For tPAL, resampling is achieved through the repetitive formation of well-defined chemical bonds (between the peptide N-terminus and the NTA-Ni), which ensures highly consistent data generation and rich event features due to the strict chemical constraints and corresponding interactions. In contrast, the nanopore reading in the other two resampling strategies is based on resistive pulse nanopore sensing. As a result, each resampling event produces highly variable outcomes, largely because the peptide is not conformationally constrained during data production, leading to significant event variability. This inconsistency can be clearly observed by comparing the results demonstrated in each resampling cycle in corresponding results (**Figure RL 1.2**).

4. **Superior spatial resolution:** Again, resampling during tPAL is achieved through the repetitive formation of well-defined chemical bonds with the peptide N-terminus. As a result, only few amino acids at the N-terminus of the peptide dominate the signal production, leading to a high spatial resolution (**Figure RL 1.3**). In contrast, the other two demonstrations report that 13-29 amino acids contribute simultaneously to each read, almost impossible for data interpretation or sequence decoding (**Figure RL 1.4**).

Alternatively, these methods often require the use of extremely long or unnatural spacer strands for demonstration.

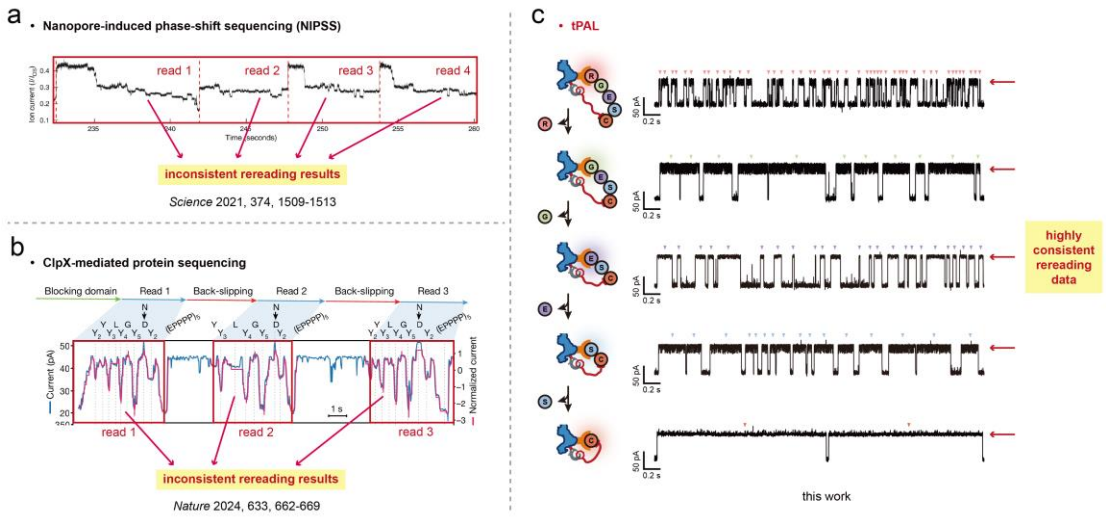


Figure RL 1.2. The technical advantage of tPAL method regarding data consistency.

5. **Resampling of the same peptide with voltage sweeping:** tPAL also enables resampling of the same molecule however at different voltages, including positive and negative voltages as well. This unique sensing modality enables the acquisition of significantly more sensing data, thereby facilitating more robust analyte identification.

It is worth noting that the NIPSS strategy (peptide-DNA conjugation) was initially developed by our group (Nano Letters, 2021, 21, 6703-6710). However, during the course of technological development, we identified the fundamental limitation in achieving sufficient resolution for practical peptide sequencing. This motivated us to explore alternative approaches and ultimately led to the development of a high resolution nanopore-based peptide analysis platform: the MspA-NTA-Ni pore and the tPAL strategy described in this manuscript.

Eventually, we highly appreciate Reviewer 1 for raising the insightful and constrictive discussions of technological comparison for resampling. The literatures you listed all represent significant milestones in the field of single-molecule nanopore protein sequencing. This tPAL method however achieves a breakthrough in both the controlled resampling and the spatial resolution. In response to Reviewer 1's inquiry, we have added detailed discussions comparing the technical and performance differences between this technique and all relevant prior works. For your convenience, these revised sections are provided below:

"At each stage, the N-terminus of the peptide can be repeatedly read, efficiently accumulating a large number of events for downstream data analysis. Compared to other nanopore-based peptide resampling strategies[16,17], tPAL performs high-frequency rereading at each cleavage state by chemical interactions, rather than repeatedly reading

the entire peptide by enzymatic activities. This resampling strategy is strictly unidirectional, and peptide backstepping is impossible. Unlike resistive pulse sensing, a tPAL event is generated only when the peptide N-terminus is chemically constrained, resulting in more consistent event features that are highly desirable for downstream sequence decoding."

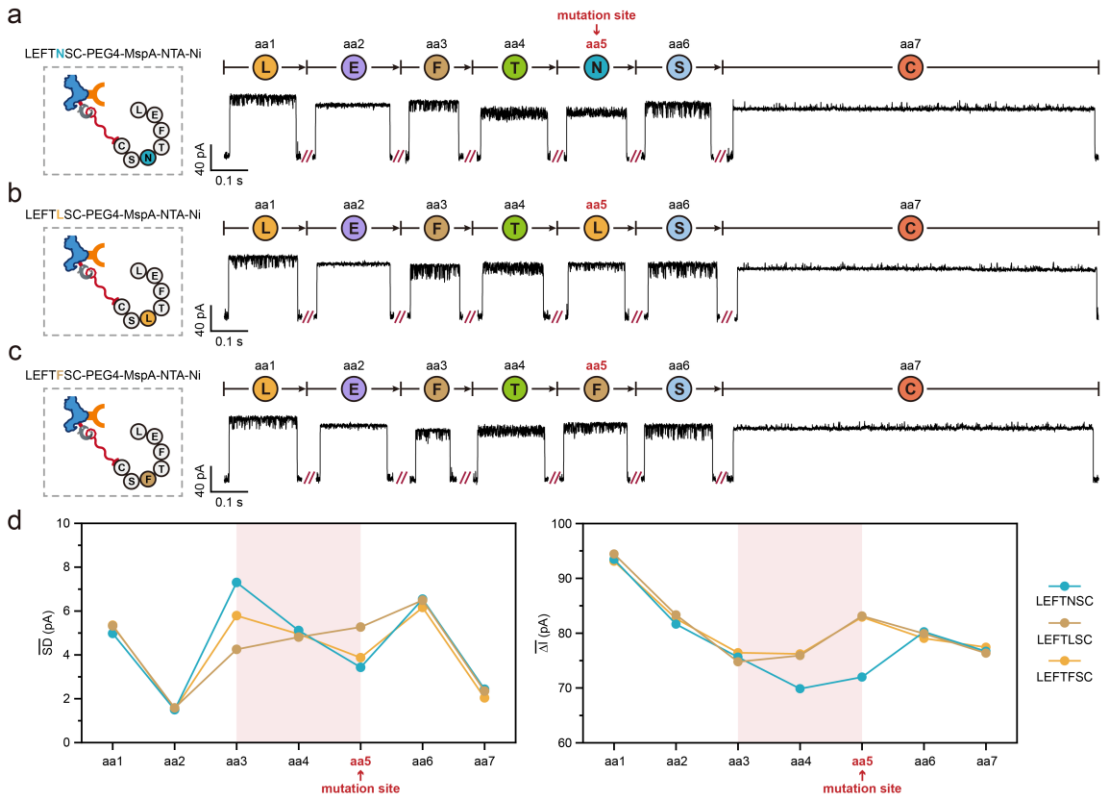


Figure RL 1.3. Sequential cleavage of peptides with single amino acid variation at the fifth position. (**a-c**) Representative events obtained by sequential cleavage of (**a**) LEFTNSC, (**b**) LEFTLSC and (**c**) LEFTFSC. The measurements were performed as described in **Methods**. For each peptide, six steps of enzymatic cleavages were observed, generating seven stages referred to as aa1-aa7, respectively. (**d**) Comparison of $\overline{SD}$ (left) and $\overline{\Delta I}$ (right) at different stages between LEFTNSC, LEFTLSC and LEFTFSC.

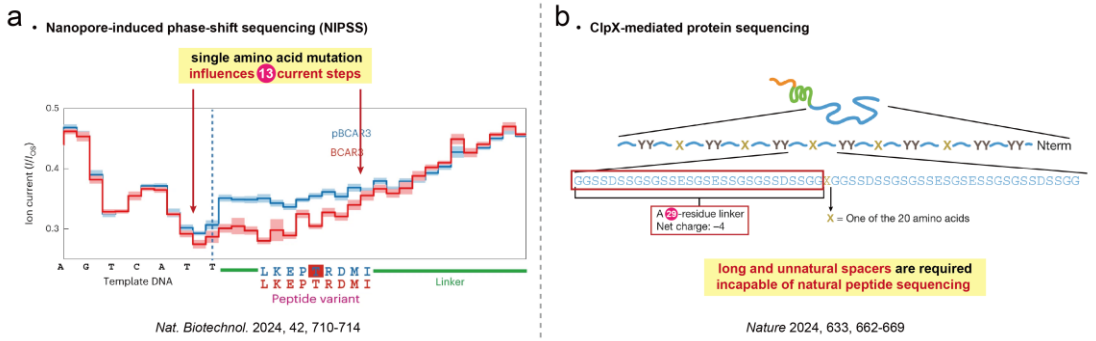


Figure RL 1.4. Limited spatial resolution of strand sequencing method. The results suggest that these two approaches, which rely on resistive pulse sensing, has a spatial resolution corresponding to 13-29 aa.

Review Comments:

Fig 1c – description in figure legend and accompanying text (98-103) requires some additional information. It would be useful to clarify what method was used to separate the different protein species – SDS-PAGE/BN-PAGE etc. or include additional relevant information in the methods.

Response:

Thanks Reviewer 1 for raising this constructive suggestion on figure legend. Strictly following your request, we have revised the figure legend of **Figure 1c** and accompanying text to explicitly describe the separation method. Additional details are also provided in the methods section of the revised manuscript as well. For your convenience, corresponding method details are pasted below:

In the figure legend:

"*Gel electrophoresis was performed using a 10 % SDS-PAGE gel (**Methods**).*"

More details provided in the **Methods** section:

"…*To separate (N90C, N93AzK)$_1$(M2)$_7$ from the mixture, gel electrophoresis of the previously collected fractions was performed using a 10% SDS-PAGE gel. The electrophoresis was run at 160 V for 16 h. The gel was stained with Coomassie blue staining solution (1.25 g Coomassie Brilliant Blue R250, 225 mL methanol, 50 mL acetic acid, 225 mL ultrapure water) for 4 h. Thereafter, the gel was de-stained with de-staining buffer (400 mL methanol, 100 mL acetic acid, 500 mL ultrapure water) until protein bands were clearly visible. The protein band corresponding to (N90C, N93AzK)$_1$(M2)$_7$ was excised from the gel, crushed, and immersed in extraction solution (150 mM NaCl, 20 mM HEPES, 0.2% DDM, 0.5% Genapol X-80, 10 mM EDTA, pH 7.5) for 12 h. The supernatant was collected and stored at -80°C for subsequent use…*"

Review Comments:

Previous publications on the topic of single-molecule peptide sensing using nanopores have substituted all 20 proteinogenic residues into a model analyte to compare the discriminative power of their respective sensing methods.

Motone, K., Kontogiorgos-Heintz, D., Wee, J., Kurihara, K., Yang, S., Roote, G., Fox, O. E., Fang, Y., Queen, M., Tolhurst, M., Cardozo, N., Jain, M., & Nivala, J. (2024). Multi-pass, single-molecule nanopore reading of long protein strands. Nature, 1-8. https://doi.org/10.1038/s41586-024-07935-7

Ouldali, H., Sarthak, K., Ensslen, T., Piguet, F., Manivet, P., Pelta, J., Behrends, J. C., Aksimentiev, A., & Oukhaled, A. (2020). Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. Nature Biotechnology, 38(2), 176-181.

https://doi.org/10.1038/s41587-019-0345-2

It would strengthen the manuscript to extend the scope of the work presented in either Figure 4 or supplementary Figure 11, including further substitutions in order to demonstrate discrimination of all 20 (or 19 without cys) proteinogenic amino acids and show that they are all separable in this sensing regime.

Response:

We sincerely thank Reviewer 1 for this constructive suggestion. We fully agree with the reviewer that demonstrating the capability to distinguish substitution of 20 proteinogenic amino acids within a single model peptide would be a strong testament to the resolution of tPAL. In response, we have undertaken a significant and comprehensive effort to address this challenge.

Specifically, we synthesized and tested 20 XTRSC peptide variants, with the mutation site located at the first residue of the N-terminus. To unambiguously discriminate all 20 peptide variants, we employed a novel recording strategy based on multi-voltage sweeping, a unique feature of tPAL that enables repeated resampling of the same peptide under different measurement conditions for an unlimited number of times, as the peptide remains anchored to the pore throughout the measurement. Experimentally, even if some peptides exhibit similar signals at a certain voltage, differences may become apparent in other voltage dimensions, thereby enabling unambiguous discrimination.

Experimentally, each peptide variant was subjected to programmed cycling voltage recordings at +60 mV, +100 mV and +140 mV (**Figure RL 1.5a**). Representative events of these peptides at different voltages are displayed in the **Supplementary Figure 17-22**. All these sensing events exhibit rich features in current fluctuation, most of which can be directly distinguished even visually. Three independent measurements were performed for each peptide to guarantee the result consistency.

Notably, CTRSC, HTRSC and PTRSC each produced two distinct types of events, which makes them even easier to be distinguished from other peptides (**Figure RL 1.5b**). CTRSC is a special case, as both its C- and N-terminal cysteine residues can conjugate with the pore. However, only C-terminal cysteine conjugation could lead to detectable signals according to the sensing principle of tPAL. In this case, we speculate that the side chain of N-terminal cysteine may also coordination with NTA-Ni, thus generating the secondary type of event.

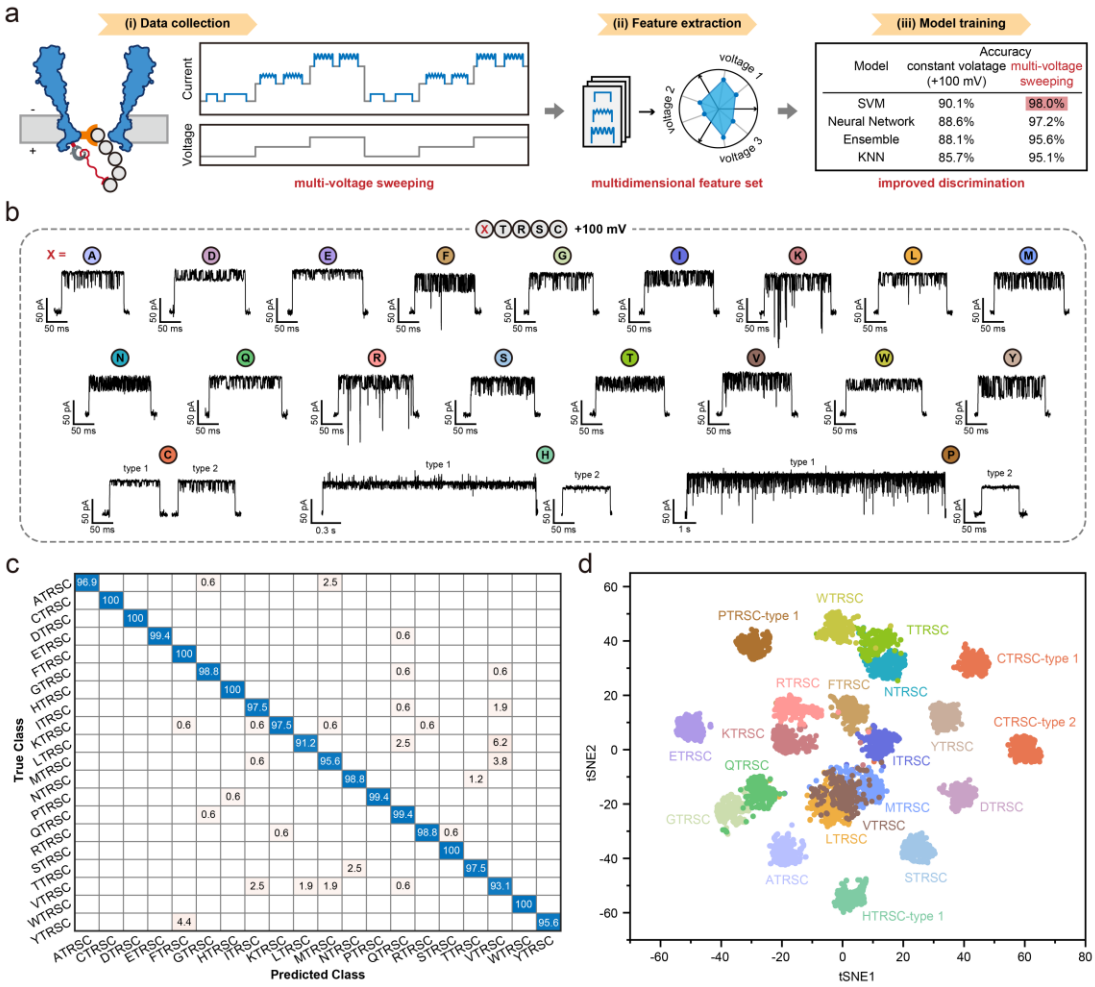


Figure RL 1.5. Discrimination of peptides with single amino acid substitutions using a multi-voltage sweeping strategy. (**a**) The workflow of multi-voltage sweeping and machine learning. In brief, each peptide was measured using a repetitive voltage protocol to acquire its tPAL events at multiple voltages (+60 mV, +100 mV and +140 mV). For each voltage, 200 events were collected per peptide to construct the dataset. For each peptide, nine features (ΔI, SD, $skew$, $kurt$, t_{off}, $median$, IQR, $spectral\ centroid$ and $spectral\ entropy$) were respectively extracted from the sensing events at each of the three voltages and subsequently concatenated to form a 27-dimensional feature matrix. Machine learning was performed with the Classification Learner toolbox of MATLAB, and model performance was evaluated by ten-fold cross-validation. With the multi-voltage sweeping strategy, the highest validation accuracy reached 98.0% with the quadratic SVM model, compared to 90.1% when using a dataset only acquired at +100 mV. (**b**) Representative tPAL sensing events of 20 XTRSC peptide variants acquired at +100 mV. Here, X represents each of the 20 proteinogenic amino acids. Specifically, CTRSC, HTRSC and PTRSC produce two types of nanopore events, defined respectively as type 1 and type 2. tPAL events of 20 XTRSC peptide variants acquired at other voltages are demonstrated in **Supplementary Figure 17-22**. (**c**) The confusion matrix of peptide classification generated by the quadratic SVM model using a dataset from multi-voltage sweeping. (**d**) t-SNE visualization of 20 XTRSC peptide variants (n=4200). The plot was generated by reducing 27 features derived from multi-voltage measurements into a two-dimensional space. For HTRSC and PTRSC, only type 1 events were collected for plotting due to the infrequent occurrence of type 2 events.

To further evaluate the discrimination capability of 20 XTRSC peptide variants, we conducted a machine learning analysis. For each peptide, three 9-dimensional feature

matrixes derived from three voltages were concatenated into a 27-dimensional feature matrix for model training. Among all classifiers, the quadratic SVM model performed best, achieving an excellent validation accuracy of 98.0% and a test accuracy of 98.2% (**Supplementary Table 7**). In contrast, when using dataset from a single voltage (either +60 mV, +100 mV, or +140 mV) for model training, the validation accuracies were significantly lower, at 92.9%, 90.1%, and 86.1%, respectively. According to these demonstrations, the advantage of the multi-voltage sweeping strategy in the resolution boosting is confirmed. For instance, the TPR of FTRSC in the training model with +100 mV dataset was only 89.4%, but with the inclusion of multi-voltage measurements, the TPR of FTRSC reached 100% (**Figure RL 1.6**).

The multi-voltage sensing data from all 20 XTRSC peptides were processed using the t-distributed stochastic neighbor embedding (t-SNE) algorithm, reducing the dimensionality from 27 to 2 for two-dimensional visualization of event distribution. The result clearly shows well-separated distributions of events corresponding to different peptides (**Figure RL 1.5d**).

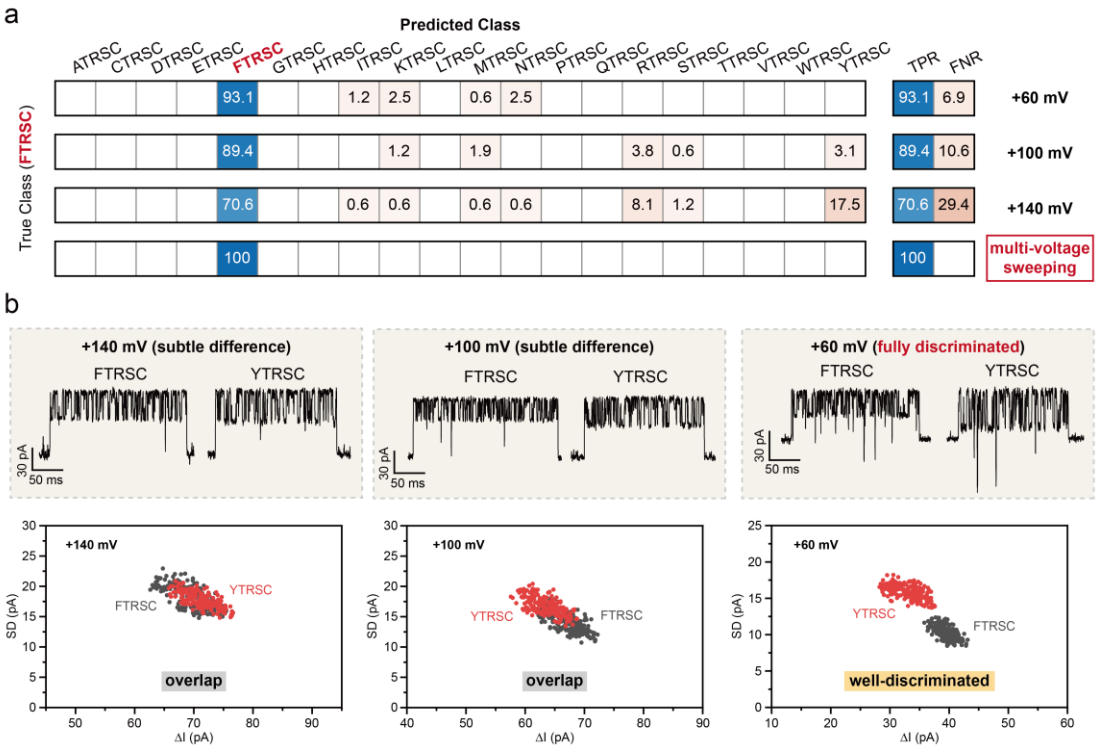


Figure RL 1.6. Machine learning performance improvement with multi-voltage sweeping strategy. (**a**) The improvement of the true positive rates (TPR) for FTRSC prediction with multi-voltage recording. (**b**) Top: Representative tPAL events acquired with FTRSC and YTRSC at +140 mV, +100 mV and +60 mV. Bottom: The scatter plots of ΔI versus SD for tPAL events acquired with FTRSC and YTRSC at +140 mV, +100 mV and +60 mV. While the event distributions have some overlap at +140 mV and +100 mV, they are well-discriminated at +60 mV.

In summary, with the implementation of multi-voltage sweeping strategy, we have successfully demonstrated the ability of tPAL to distinguish all 20 amino acid mutations in the same model peptide. However, to ensure clarity, all other experiments in the

manuscript only present data acquired at a single voltage (+100 mV). We sincerely appreciate reviewer's constructive suggestion, which has allowed us to have the opportunity to show the multi-voltage sweeping mode of sensing, which may be widely applied in the analysis of more complicated nanopore peptide sequencing data in prospects.

Eventually, to fully address this extremely constructive comment, all above described results along with relevant discussions were added in the revised manuscript. For the ease of your reference, a fraction of the added discussion is included below:

"After establishing the tPAL methodology, its sensitivity toward single amino acid mutations was further investigated. Twenty XTRSC peptide variants, with mutations introduced at their N-terminus, were tested. Representative events of these peptides obtained at +100 mV are shown in ***Extended Data Figure 1b****. All events exhibit rich characteristics, most of which can be directly distinguished by visual observation. … The results clearly demonstrate the high resolution of tPAL technology and the effectiveness of multi-voltage sweeping. However, for the sake of clarity, all subsequent experiments only present data acquired at +100 mV, unless otherwise stated."*

Review Comments:

Features of events were shown to be reproducible across replicates (supplementary tables 5 & 10). Error bars are included in all graphs where appropriate. The authors report a Cubic SVM model (supplementary table 6 & supplementary Figure 18) able to discriminate a peptide and peptides representing stepwise cleavage products to high accuracy (99.5%). Both test accuracies and those from cross-validation are reported, along with the results from a variety of other models. In the legend for supplementary table 6 it is stated that datasets of 400 events were generated for each peptide; however, it is not clear how many nanopores this came from. The peptides discriminated by the Cubic SVM model vary in both length and N-terminal amino acid, thus I think some clarification is needed for the following wording: "illustrating the tPAL strategy's efficacy in distinguishing single amino acid differences within peptide sequences" (240-242).

Response:

Thanks Reviewer 1 for providing this rigorous comment on the provision of more experimental details. Your recognition of our data reproducibility and high classification accuracy is sincerely appreciated. In order to ensure the reproducibility, the statistics of each peptide in the machine learning dataset were randomly selected from three completely independent measurements using different pores. We apology for not disclosing sufficient technical details in the original manuscript. To fully address this, we have revised the legend of original **Supplementary Table 6** (renumbered as **Supplementary Table 9** in revision), strictly following your request.

We also apology for the ambiguity in the original lines 240-242. The machine learning

model here demonstrates high classification accuracy of stepwise shortened peptides, with each pair differing by only a single amino acid. Therefore, it is concluded that tPAL strategy can effectively distinguish peptides with single amino acid difference. Following your suggestion, we have also revised this sentence to avoid ambiguity. For the ease of your reference, the revised sentences are pasted below:

"A dataset was constructed by randomly selecting 400 events from three independent measurements of each peptide-PEG4-MspA-NTA-Ni variant"

"illustrating the tPAL strategy's efficacy in distinguishing stepwise shortened peptides"

Review Comments:

It would be useful to clarify the statement "Additionally, for *de novo* peptide sequencing, a reference database and corresponding AI-driven algorithms are essential for decode nanopore events into the corresponding peptide sequences." (414-416) The use of a reference database isn't sequencing *de novo*.

Response:

We deeply appreciate Reviewer 1 for this rigorous comment. We sincerely apologize for the confusion caused by imprecise wording in the original manuscript. We wish to clarify that the term "reference database" mentioned in the original text was intended to refer to an annotated dataset used for the supervised training of machine learning model, rather than a reference protein database used for complete sequence matching.

Traditionally, *de novo* peptide sequencing refers to the determination of peptide sequence directly from experimental data without reliance on protein database or genomic information for sequence alignment (Expert Review of Proteomics, 2011, 8, 645-657; Analytica Chimica Acta, 2023, 1268, 341330). Based on the direct correlation between sensing events variations during enzymatic cleavage and the corresponding peptide sequence, tPAL method still holds great promise for *de novo* peptide sequencing, especially with the newly included results demonstrating decoding of tPAL data in the revision.

Specifically, we constructed a training dataset of nanopore events corresponding to certain known k-mers to train the machine learning model (**Figure RL 1.7**). This trained model could be employed to recognize the k-mer identity of different event stages during peptide sequential cleavage (**Figure RL 1.8**). The full peptide sequence is reconstructed by aligning the overlapping regions among consecutive k-mers (**Figure RL 1.9**). With the continued expansion of k-mer databases in the future, tPAL method is expected to achieve *de novo* decoding of unknown peptide sequences.

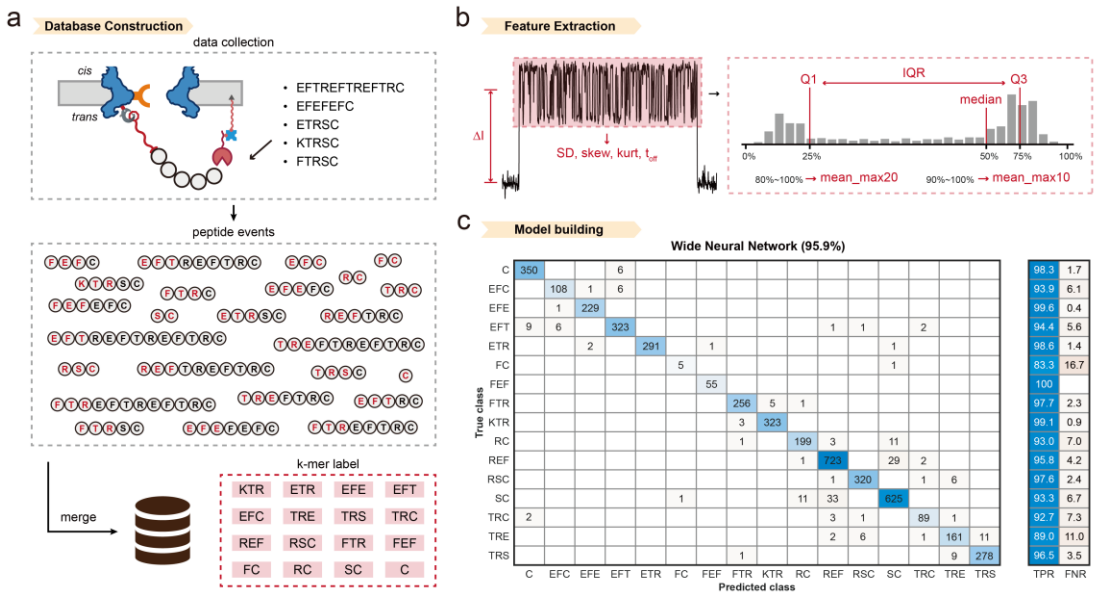


Figure RL 1.7. Construction of k-mer database and machine learning. (**a**) Database construction. Peptide events acquired during sequential cleavage of EFTREFTREFTRC, EFEFEFC, ETRSC, KTRSC and FTRSC were collected and merged to form the database. Peptide events with identical sequences in their first three residues are assigned the same 3-mer label, while events from dipeptides and amino acid are assigned corresponding 2-mer and 1-mer labels. (**b**) Feature extraction. Eleven event features were extracted from events to form a feature matrix, including ΔI, SD, $skew$, $kurt$, t_{off}, $median$, IQR, $Q1$, $Q3$, $mean\_max10$, $mean\_max20$. Here, $mean\_max10$ is defined as the mean value of top 10 percent of the current blockage ranked from largest to smallest, and $mean\_max20$ is defined as the mean value of top 20 percent of the current blockage ranked from largest to smallest. (**c**) The confusion matrix generated by the wide neural network model, which achieves a validation accuracy of 95.9%.

We again thank Reviewer 1 for your helpful comment regarding the technical description. We fully recognized that the phrase “reference database” in the original statement could be misleading in this context. Accordingly, we have removed the original sentence to avoid ambiguity and replaced it with a more detailed and precise description of our sequence decoding process. For the reviewer’s convenience, we have pasted the relevant discussions below:

“Based on the insight that peptide events are predominantly influenced by the first few amino acids at the N-terminus, it is in principle possible to directly decode peptide sequence from tPAL results acquired during sequential enzymatic cleavage. … … Though demonstrated here using a miniaturized k-mer database constructed from events collected with existing peptides, further expansion of the k-mer database and optimization of machine learning algorithms hold considerable promise for practical sequencing applications.”

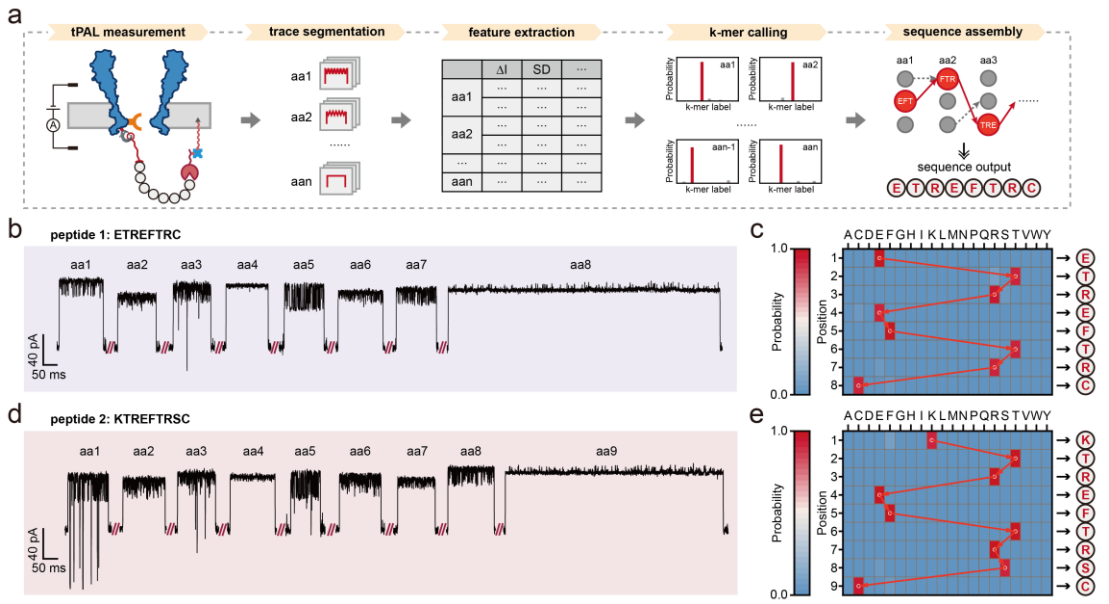


Figure RL 1.8. Peptide sequence decoding from tPAL events. (**a**) The workflow of machine learning-assisted sequence analysis. Briefly, the raw data obtained from peptide sequential cleavage is first divided into different stages. Feature extraction is then performed for sensing events within each stage to generate the feature matrix. Subsequently, a machine learning model pre-trained with a k-mer dataset is employed for event prediction, yielding k-mer labels and probability distributions for each stage. Finally, sequence assembly is conducted based on sequence overlap between k-mers from adjacent stages, outputting the complete peptide sequence. (**b, d**) Representative events acquired with (**b**) ETREFTRC and (**d**) KTREFTRSC during sequential cleavage. The measurements were performed as described in **Methods**. (**c, e**) Amino acid probability heatmaps derived from (**c**) ETREFTRC and (**e**) KTREFTRSC measurements. The probabilities of 20 amino acids at each position are calculated by normalizing the frequencies of all k-mers covering that position in prediction results. The optimal path (red polyline) connects the amino acid with the highest probability at each position, forming the predicted peptide sequence on the right side of each heatmap.

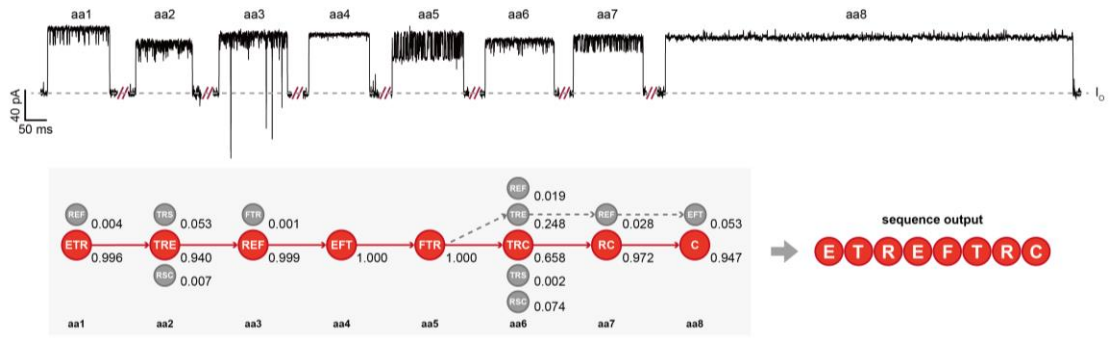


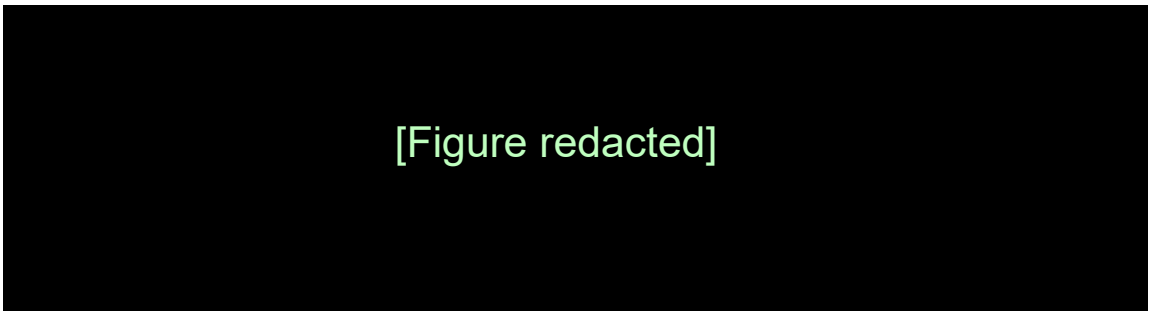
Figure RL 1.9. tPAL measurements of ETREFTRC and sequence assembly. Top: Representative traces acquired with ETREFTRC during sequential cleavage. Seven enzymatic cleavages were identified during the sequential cleavage of ETREFTRC, leading to eight distinct stages. Bottom: Sequence assembly graph. Each node in the graph represents a candidate k-mer predicted from the sensing events at corresponding stage, with their probability annotated alongside. Directed edges represent valid connections in which the suffix of a preceding k-mer exactly matches the prefix of the current k-mer. The optimal path, highlighted in red, is defined as the route exhibiting the highest cumulative probability while traversing exactly one node per stage. All remaining nodes and suboptimal paths are depicted in gray. The final peptide sequence was derived by sequentially merging the k-mers along this optimal path.

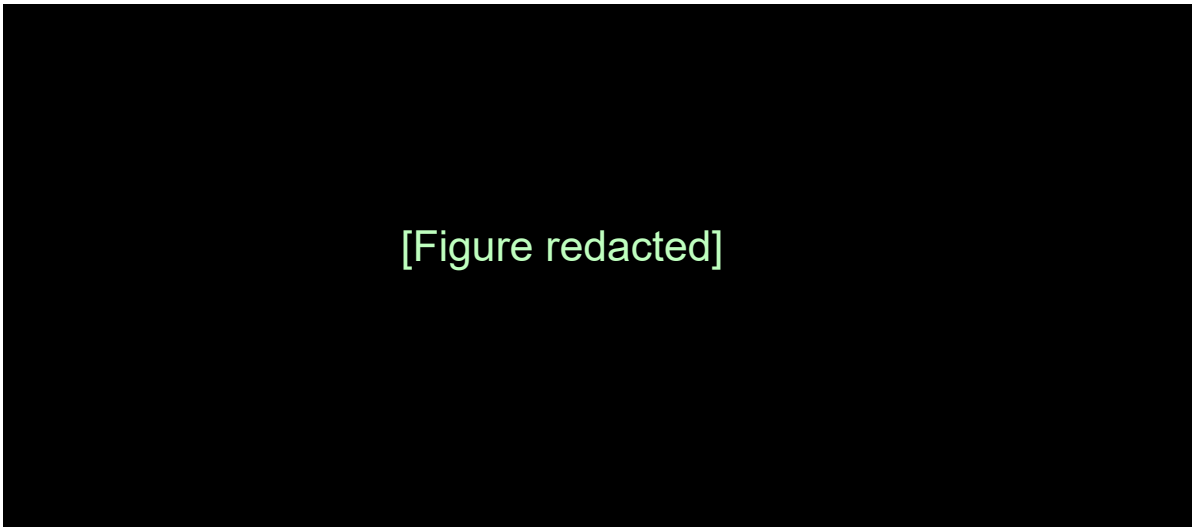
Review Comments:

As the tPAL approach is currently limited to one peptide per experiment (of a given length), it would also be useful for the authors to comment on how this limitation could be overcome both with respect to generating training data for machine learning but also for enabling the method to address biological questions.

Response:

We sincerely appreciate Reviewer 1's insightful and thoughtful comments regarding the current limitation in throughput. However, we respectfully argue that the priority at this stage is to demonstrate conceptual advances in addressing the more urgent challenges of sequencing resolution, with further engineering efforts to improve throughput being pursued in later stages. To ensure high-quality data demonstration, the target peptide is covalently linked to the pore, and measurements were performed using a single nanopore at a time, appearing that this technique is insufficient in the measurement throughput.





However, the development of such a system requires extensive exploration and interdisciplinary collaboration, which is beyond the scope of the current study, which serves to demonstrate the technical breakthrough. To fully address this issue, we have incorporated a discussion of potential solutions to throughput issue in the revised manuscript. The relevant content is included below for your convenience:

"The integration with high-density nanopore arrays has the potential to significantly increase the measurement throughput."

Review Comments:

Overall, this manuscript makes a major conceptual and technical contribution to the field of nanopore protein sequencing by introducing a new method for peptide analysis. I strongly support publication after minor clarifications and extensions as noted above.

Response:

We sincerely thank reviewer 1 for the thorough and constructive evaluation of our manuscript. Your recognition of this work's significance to the field is greatly appreciated. In this round of revision, we have carefully addressed all the specific points raised by the reviewer. Specifically, we have incorporated additional data, clarified experimental details and technical descriptions, and added discussions on potential avenues for future optimization. We are confident that the revised version now presents a more rigorous, comprehensive, and accessible account of our research. We are grateful for the opportunity to improve our work and thanks again for your support to the publication of this study.

Reviewer #2:

Review Comments:

Chen and colleagues describe a method for peptide identification using nanopores. The system relies on a heterooctameric MspA nanopore in which one protomer carries two modification sites: the first attaches a Ni-NTA moiety, and the second attaches a maleimide moiety. The latter enables the peptide to be tethered to the nanopore, while the former allows recognizing to the N-terminus of the attached peptide. Each tested peptide yields a characteristic current blockade, which is likely to be strongly influenced by the identity of the terminal amino acid. An aminopeptidase anchored to the membrane then cleaves the last amino acid, providing a new peptide's recognition by the nanopore. This process is repeated until all amino acids are removed. This approach is an ingenious extension of prior work from the Huang group. The manuscript is well written, and the data are convincing. However, I have reservations about the potential of this method to deliver a practical sequencing technology.

Response:

We sincerely thank Reviewer 2 for your thoughtful evaluation of our manuscript and for your accurate description of the core innovation of this technique. We are particularly grateful for your appreciation of the ingenuity of our method, the clarity of the writing, and the convincing quality of the data. Notably, this method features so far the highest spatial resolution for nanopore peptide sequence information decoding, enabling the identification of single amino acid mutations, post-translational modifications (including methionine oxidation, tyrosine phosphorylation, lysine acetylation, and lysine trimethylation), and the insertion of unnatural amino acids within the peptide. In principle, it should also be suitable for nanopore sequencing of chiral peptides (**Extended Data Figure 4**) when compatible mirror-image enzymes could be employed, suggesting a high level of generality and a promising future for this technique.

In this round of revision, following the suggestions of all reviewers, we have further included additional direct peptide sequencing results, demonstrating direct sequencing of peptide of repeating sequence and direct decoding of sequencing results. We believe that they now demonstrate a prototype of nanopore peptide sequencing that no previous nanopore technologies have achieved. Here, this unique tPAL configuration has enabled a significantly higher spatial resolution, broad compatibility with various peptide charges, and strictly prohibits backstepping/backsliding of peptide sequence reading.

First, different from the strand sequencing, tPAL is based on well-defined and reversible chemical interactions, which enhances recognition fidelity by reporting extremely rich event features caused by well-defined chemical interactions. Existing strand sequencing strategies however requires long peptide spacers (**Figure RL 2.1**) and fails to truly decode the peptide sequence.

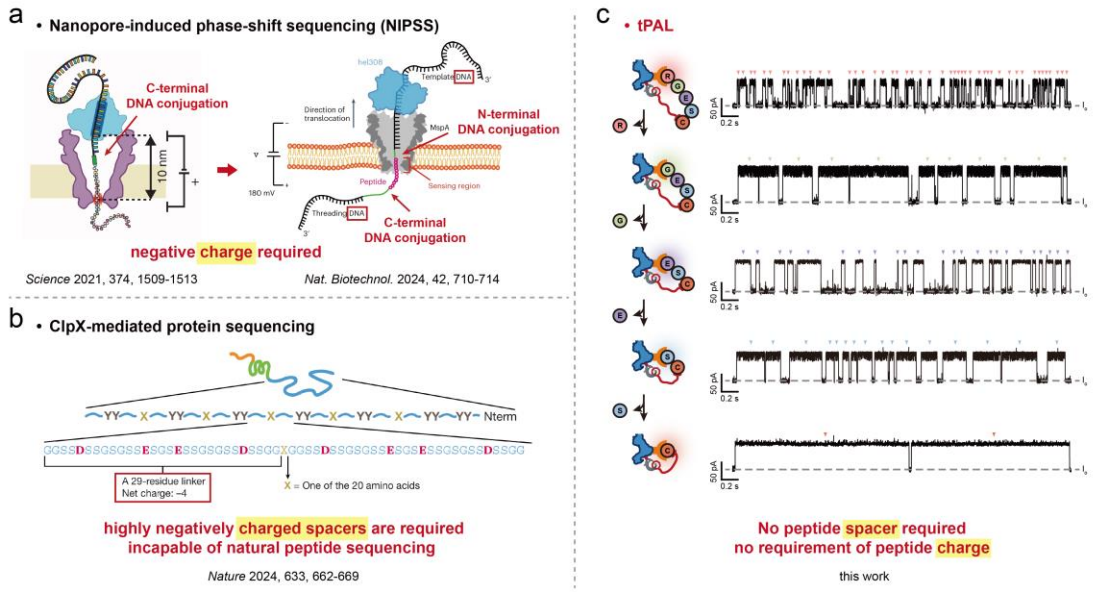


Figure RL 2.1. The substrate limitation of strand sequencing methods. Existing strand sequencing strategies require peptide substrates or spacers with high negative charges. In contrast, tPAL method is capable of detecting peptides with different charges, and does not require any spacers.

Second, the tPAL configuration also ensures strict and unidirectional single amino acid stepping by enzymatic cleavage, thereby preventing the unpredictable strand backsliding/backstepping that frequently occurs in strand sequencing. Based on my in person discussion with Dr. Keisuke Motone (the first author of the manuscript, *Nature* 2024, 633, 662-669) during a recent nanopore conference held in Tokyo, the backsliding issue is severely preventing the consistency of the measurement and is prohibiting the high fidelity decoding the corresponding peptide sequence. However, with this tPAL configuration, strand backsliding is technically impossible, enabling extremely high consistency of data production. I speculate that the backsliding issue (**Figure RL 2.2**) in the strand sequencing approaches should result from the lack of charges of the target peptide strands (**Figure RL 2.1a, b**). It is for this reason, highly negative charged spacer (**Figure RL 2.1**) peptide sequences were required in their test. However, with tPAL, it is compatible with any peptide charge or sequence, which is more practical for nanopore sequencing of real peptide.

What's more, the unique configuration of tPAL also allows for repetitive measurement of the same peptide as many times as needed, further enhancing the confidence in sequence identification.

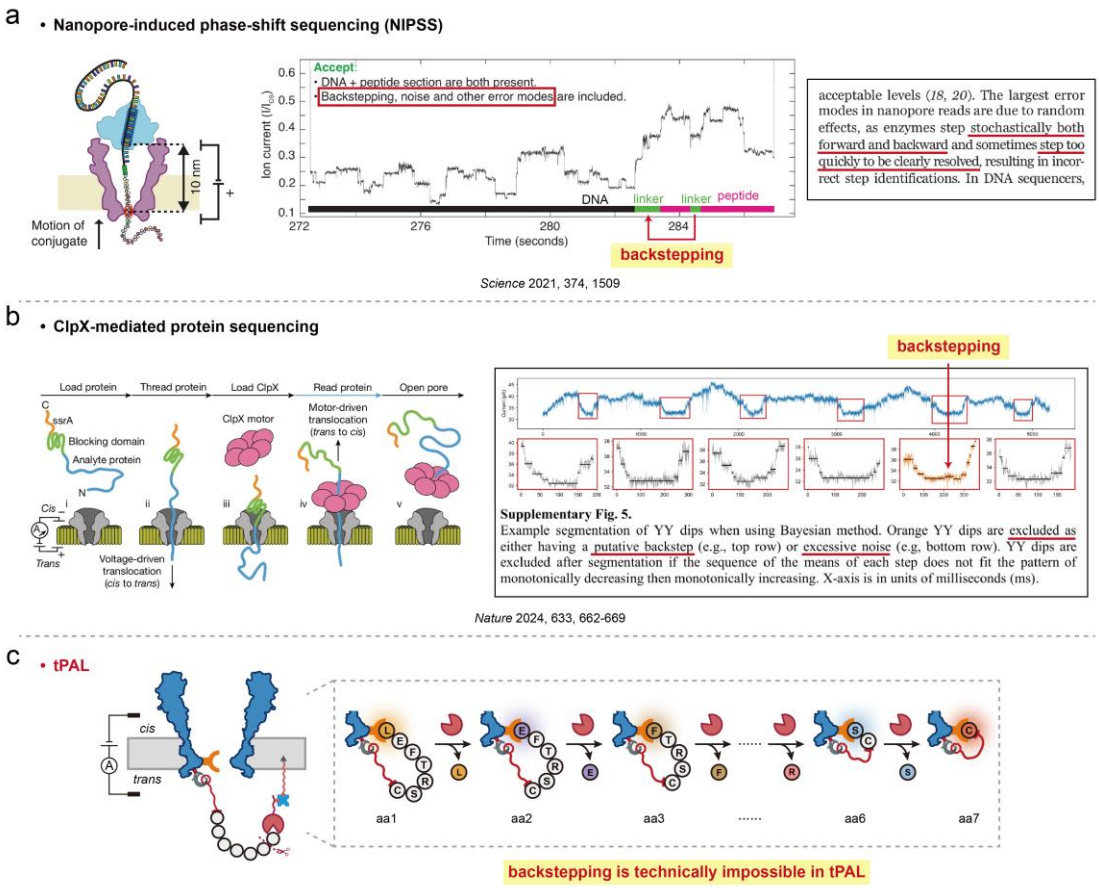


Figure RL 2.2. The technical advantage of tPAL method regarding unidirectional single amino acid stepping. Strand sequencing methods, such as NIPSS and ClpX-mediated protein sequencing, often face challenges related to strand backsliding. In contrast, the tPAL method demonstrates a notable advantage in achieving unidirectional single amino acid stepping through stepwise enzymatic digestion.

In the later review comment, the reviewer also mentioned the need for peptide C-terminus conjugation. Actually, with the peptide-DNA conjugation strategy, it even requires the target peptide to be sandwiched between two DNA strands, thus requiring simultaneous and orthogonal modification at both the N- and C-termini of the peptide (**Figure RL 2.3a**). Similarly, the ClpX strategy also necessitates the conjugation of specific peptide sequences to the target peptide (**Figure RL 2.3a**). However, with tPAL, modification is only required around the vicinity of the C-terminus (**Figure RL 2.3b**) not even necessary at the C-terminus, making this approach even simpler, or in other words, much more practical.

Here, tPAL, which offers distinct technical advantages over all existing nanopore-based peptide sequencing methods, is designed to address the current challenges in nanopore protein sequencing. As a result, it is actually more practical than all existing nanopore peptide sequencing strategies, as it does not require a specific peptide charge, backstepping cannot occur, and no spacer peptide is needed. It also requires labeling of only one terminal, and the label does not even need to be precisely at the terminal. In contrast, all previous nanopore peptide technologies require labeling of both termini. The publication of this work has the potential to benefit the nanopore field by drawing greater

attention to the promise of nanopore technology in proteomics. It is also expected to stimulate further research into protein conjugation chemistry, the development of nanopore-related hardware, and the advancement of associated bioinformatics algorithms. However, these outcomes can only be realized if the work is brought to the attention of the academic community. Your support would be extremely valuable to us.

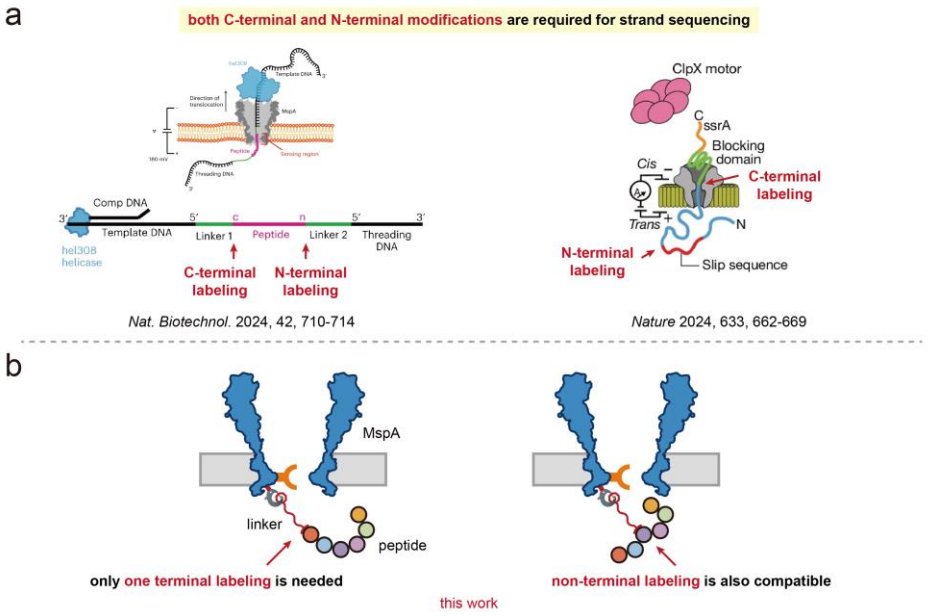


Figure RL 2.3. Comparison of peptide modification requirements between different methods. Strand sequencing methods, such as NIPSS and ClpX-mediated protein sequencing, require both C-terminal and N-terminal modifications of peptide before detection. In contrast, tPAL method is simpler in peptide processing and only requires C-end functionalization.

We are also deeply grateful for so many inspiring suggestions provided by the reviewer, which have greatly stimulated the further development of the tPAL technique. In this revised version of the manuscript, we have included numerous new results, such as tPAL sequencing of long and repetitive sequences, as well as algorithms for database construction and direct sequence decoding. We believe that the technique is now further improved compared to the original manuscript, and your insightful comments are indispensable for this improvement and are highly acknowledged.

BTW, we have also included a video clip (**Supplementary Video 1**) to provide all reviewers a live demonstration of nanopore reading of a peptide sequence all the way to the C-terminal cysteine. We trust that our dedicated work will ultimately draw your interest and reinforce your confidence in the applicability of this technique.

Given that the tPAL strategy is still in its infancy stage, we envision that numerous engineering efforts, such as the use of large nanopore arrays, enzyme engineering, and peptide terminal modification techniques, will further enhance its performance. We foresee the potential for collaborations among various research groups to ultimately refine and

perfect tPAL. However, the immediate priority is to disclose the technique to the field as soon as possible. We once again express our heartfelt gratitude to the reviewer for your support and recognition of our work. Detailed point by point responses to your comments are provided below for your convenience.

Review Comments:

Attachment to the nanopore is required for each peptide. Currently, this is accomplished using cysteine residues. This is problematic because relatively few peptides contain cysteine, and if the cysteine is not at the C-terminus, the reading is incomplete. The authors claim that nanopores can be engineered to react with other amino acids, but such chemistries exist for only a limited subset of residues. Moreover, signals from residues extending from the attachment point toward the C-terminus would still be unobservable. The only plausible solution I can envision is to develop a chemistry that enables selective reaction at the peptide's C-terminus with the nanopore; however, this remains an unresolved chemical challenge.

Response:

We sincerely acknowledge Reviewer 2 for raising the insightful discussions of peptide-nanopore coupling. We agree with the reviewer that the development of a universal, peptide C-terminus-nanopore conjugation technology would be desired for the application of tPAL. However, this challenge is not unique to this approach but is common in the whole field of single-molecule protein sequencing (nanopore and fluorescence). Emerging sequencing technologies, such as fluorosequencing (Nature Biotechnology, 2018, 36, 1076-1082) and real-time dynamic single-molecule protein sequencing (Science, 2022, 378, 186-192), also require an efficient C-terminal bioconjugation method for peptide library preparation (**Figure RL 2.4**). What's more, nanopore-based strand sequencing methods, including NIPSS (Science, 2021, 374, 1509-1513) and ClpX method (Nature, 2024, 633, 662-669), require orthogonal modifications at both the N- and C terminus (**Figure RL 2.3**), which is even more complex than tPAL.

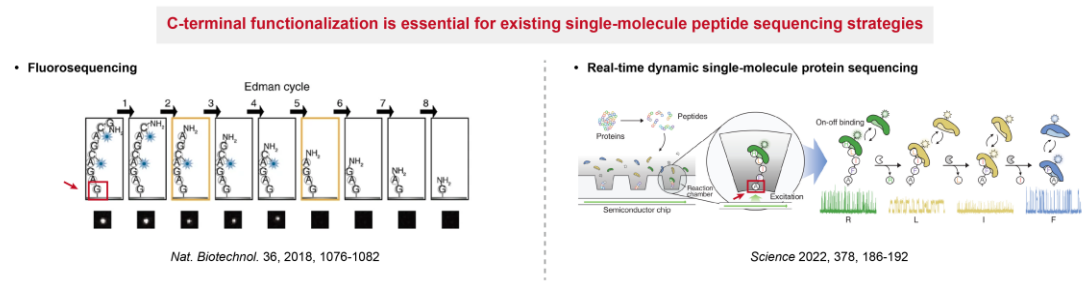
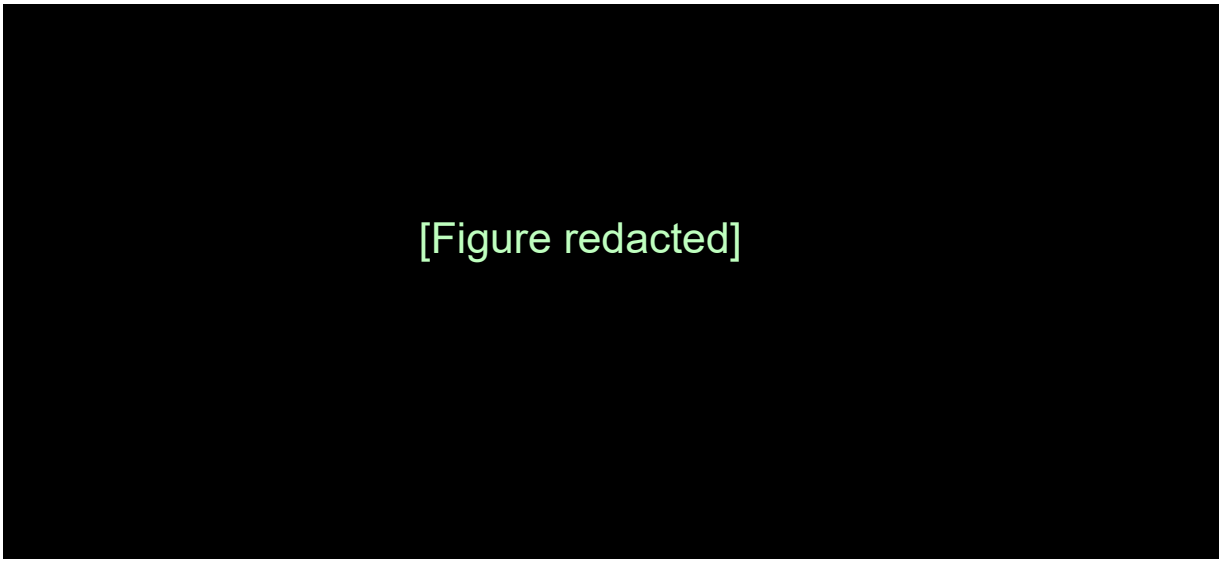
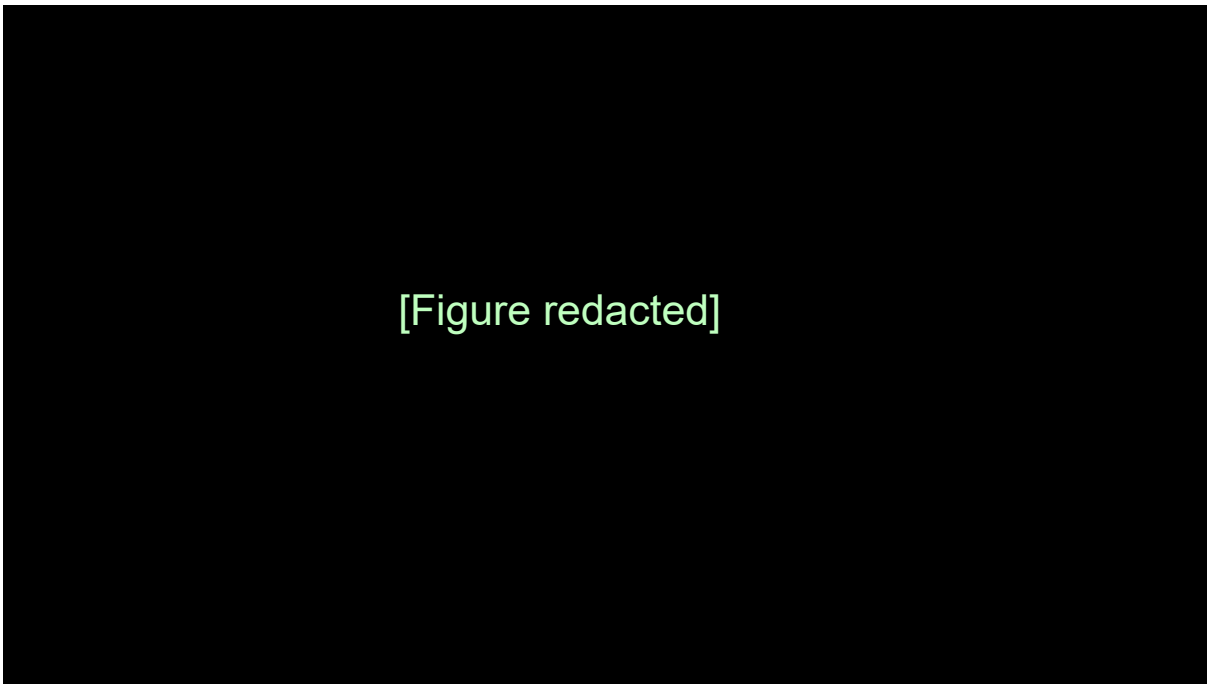


Figure RL 2.4. The urgent need for C-terminal functionalization methods in existing single-molecule peptide sequencing strategies.

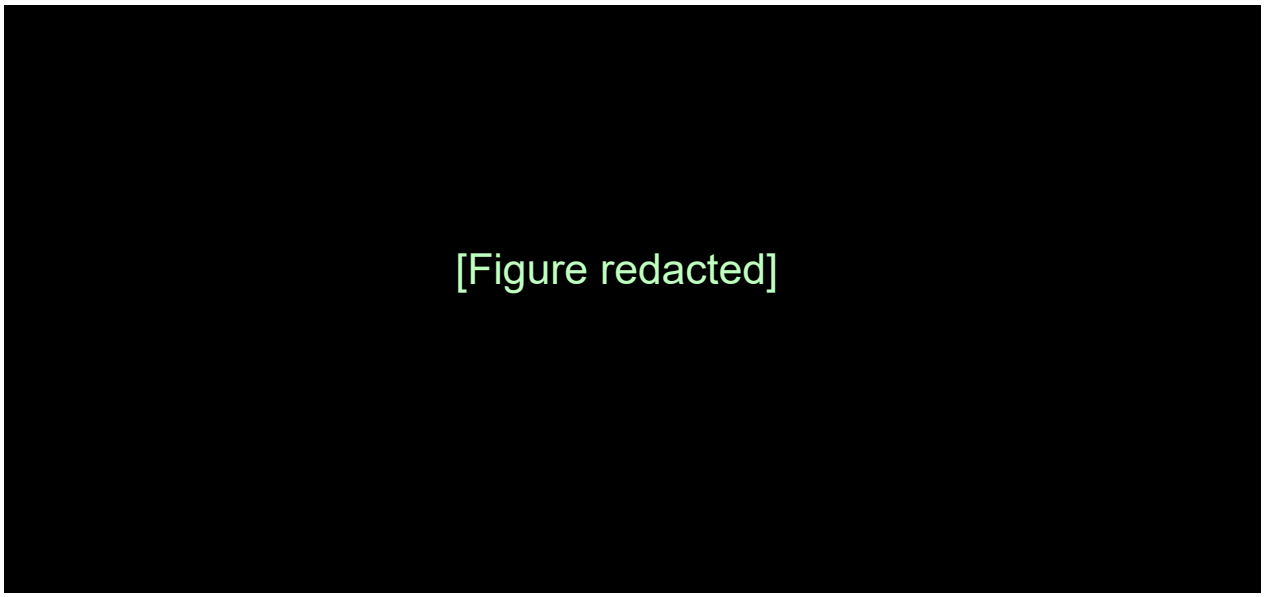
Addressing this challenge requires collaborative efforts from the broader scientific

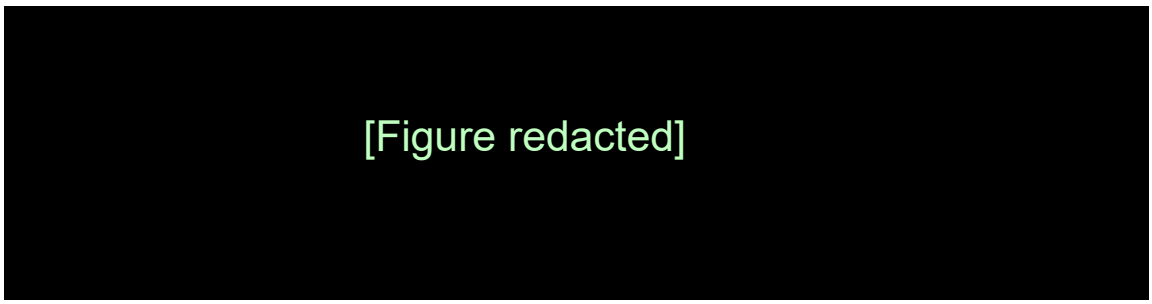
community. Fortunately, recent advances in chemical biology have provided promising solutions. Several photocatalytic methods for C-terminal carboxyl group modification have been developed (Nature Chemistry, 2018, 10, 205-211; Angewandte Chemie International Edition, 2019, 58, 8182-8186; Chemical Science, 2021, 12, 2467-2473), enabling selective introduction of alkyne group to the peptide C-terminus (**Figure RL 2.5**). These C-terminally alkynylated peptides can then be conjugated to the pore via click reaction with the azido group on the pore. These chemistries are in principle compatible with our platform and could be integrated into future iterations of our technology to enable the detection of native peptides. Notably, this photocatalytic C-terminal alkylation approach has already been adopted in fluorosequencing (ACS Chemical Biology, 2021, 16, 2595-2603), fully demonstrating its feasibility for site-selective peptide modification.





[redacted]





Eventually, we appreciate Reviewer 2 for your constructive comments, which has sparked meaningful discussions regarding the improvement to tPAL method. To address your concerns, we have included a rather comprehensive discussion on peptide terminal labeling in the revised manuscript. For your convenience, the revised sections are provided below:

"Through photocatalytic C-terminal-specific labeling methods[66-68], peptide fragments derived from natural proteins can be functionalized to enable their attachment to the nanopore. Furthermore, specific endopeptidases such as Lys-C and Arg-C can be employed to hydrolyze proteins into peptide fragments with defined C-terminal residues. By utilizing selective functionalization methods targeting the side chains of these specific C-terminal residues[69,70], these peptide fragments can also be immobilized on the nanopore for tPAL measurements."

Review Comments:

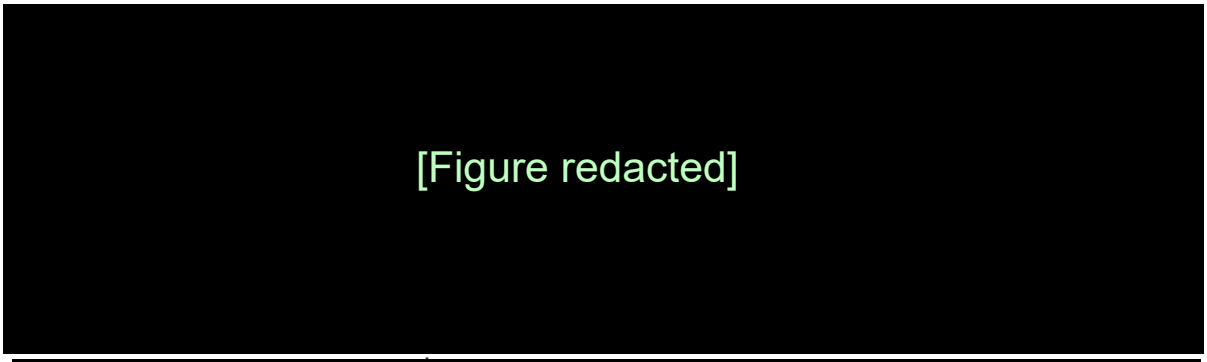
Throughput is limited because only one peptide can be analyzed per nanopore at a time. This would necessitate forming and breaking lipid bilayers after each measurement, leading to very low throughput.

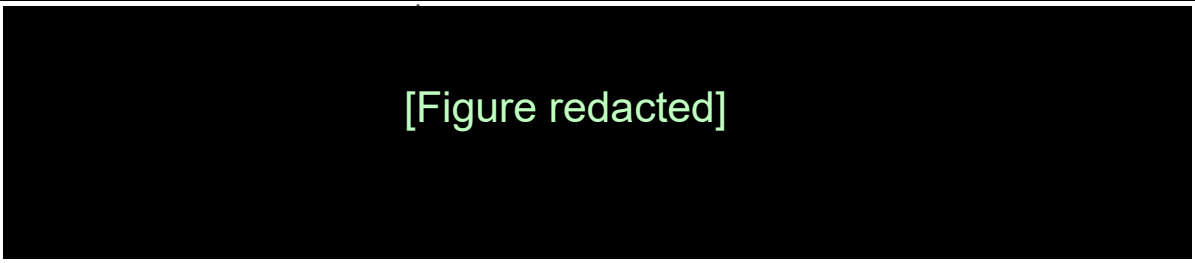
Response:

We sincerely appreciate Reviewer 2 for raising this insightful comment regarding the throughput. We fully acknowledge that the current setup exhibits limited throughput due to its single-pore, single-molecule detection nature, and the efficiency of data acquisition at its current form appear relatively low. However, we wish to clarify that, as the first demonstration of tPAL, the primary objective of this work is to establish a proof of concept for this novel peptide sequencing method. Much like the early developmental stages of nanopore DNA sequencing, initial efforts focused on validating the core principle before moving on to technical optimization and throughput scaling, as later exemplified by Oxford Nanopore Technologies during the commercialization phase.

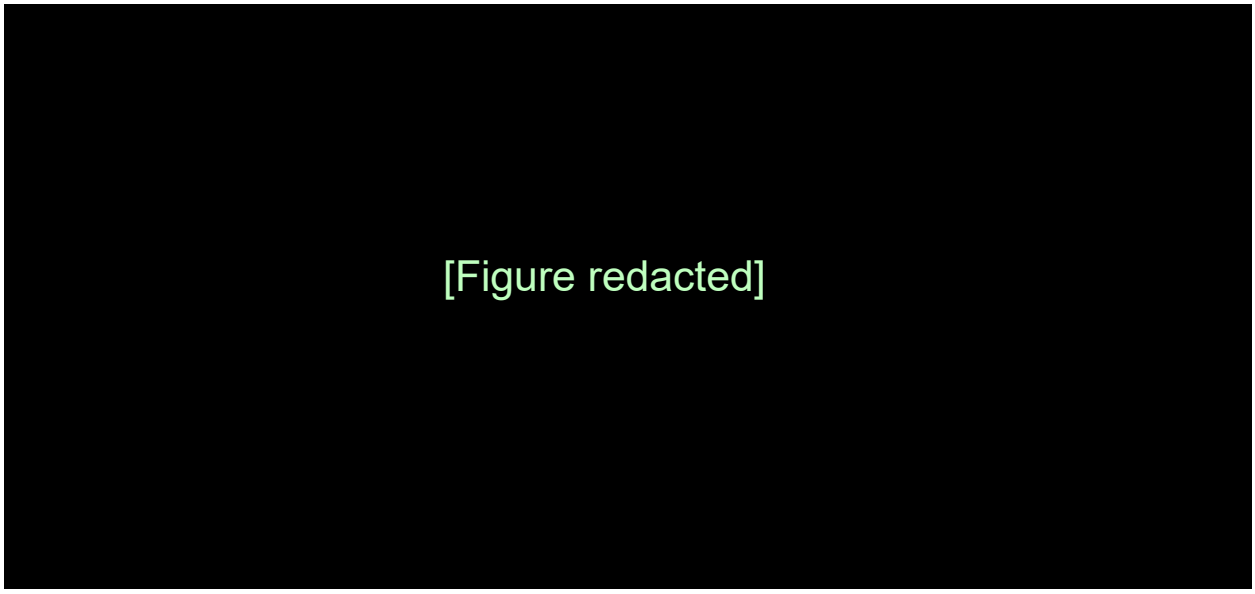
In response to the reviewer's concern about the need to reconstruct lipid bilayer after each measurement, [redacted]

[redacted] According to literature, considerable research efforts are currently being devoted to the development of such reversible chemical systems (Angewandte Chemie International Edition, 2014, 53, 8784-8787; Polymer Chemistry, 2017 8, 4038-4042; Journal of the American Chemical Society, 2017, 139, 6146-6151).





[redacted]



Nevertheless, the development of such a system requires extensive exploration and interdisciplinary collaboration, which is beyond the scope of this current study. We will actively pursue this direction in our ongoing and future work. Possibly we may collaborate with companies such as Oxford Nanopore Technologies, Portal or BGI to speed up this engineering process. Nevertheless, in the revised manuscript, we have expanded our discussions on the potential strategies for throughput enhancement. They are also pasted below for your convenience:

"The integration with high-density nanopore arrays has the potential to significantly increase the measurement throughput."

Review Comments:

The method does not provide straightforward peptide sequencing. Prior work suggests that more than ten amino acids can influence the signal, so it is highly likely that the observed blockade results from multiple residues occupying the recognition site, rather than solely the terminal amino acid. While the authors show that terminal histidine and proline can generate distinctive signals, other amino acids yield current blockades that do not appear specific to the last residue. The noise of the blockade may carry information about the entire peptide, and the kinetics of amino acid cleavage could correlate with the last few residues. Thus, the overall ionic signature as the peptide transitions from full length to fully digested may be unique to a given peptide, but this has not been demonstrated. Nonetheless, given the existence of millions of peptides in the human proteome, the required training set to address all possible sequences is likely to be enormous. To begin addressing this challenge, the authors should undertake systematic testing of peptides. One possible approach would be to examine sets that share the first two residues (e.g., LR) but differ at the third position (i.e., 20 LRX peptides). If the current delta is the same across all 20 peptides, the signal might reflect the L→R transition; subsequent experiments could then explore other residue combinations.

Response:

We sincerely thank Reviewer 2 for the constructive comments regarding the interpretability of tPAL data. In the revised manuscript, we have systematically conducted a series of experiments to evaluate the spatial resolution of the tPAL method. Specifically, we tested three peptides that differ only at the fifth position: LEFTNSC, LEFTLSC, and LEFTFSC. During real-time enzymatic cleavage, all three peptides undergo seven distinct stages, denoted as aa1 to aa7, respectively (**Figure RL 2.12**). Notably, while the sensing events at the aa3-aa5 stages showed distinct features among the three peptides, the events at the aa1 and aa2 stages were highly similar. This finding indicates that the detected events are primarily influenced by the first three N-terminal residues, with only minor contributions from later residues.

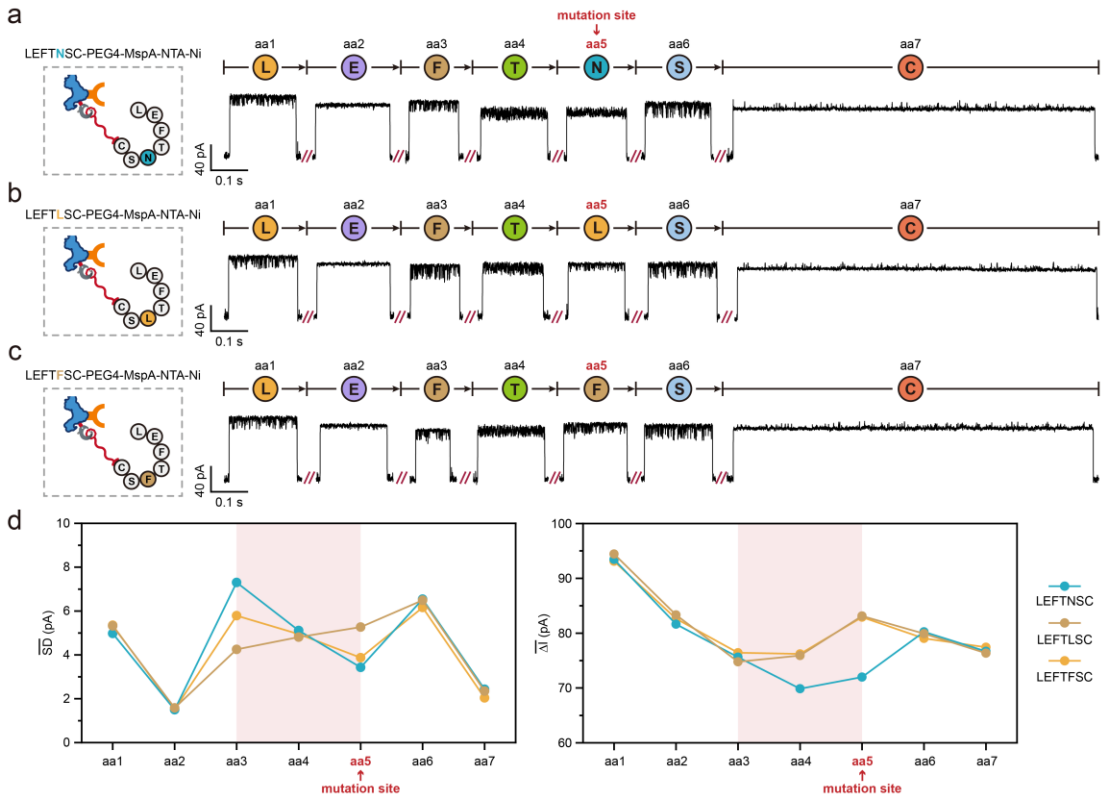


Figure RL 2.12. Sequential cleavage of peptides with single amino acid variation at the fifth position. (**a-c**) Representative events obtained by sequential cleavage of (**a**) LEFTNSC, (**b**) LEFTLSC and (**c**) LEFTFSC. The measurements were performed as described in **Methods**. For each peptide, six steps of enzymatic cleavages were observed, generating seven stages referred to as aa1-aa7, respectively. (**d**) Comparison of $\overline{SD}$ (left) and $\overline{\Delta I}$ (right) at different stages between LEFTNSC, LEFTLSC and LEFTFSC.

Furthermore, to demonstrate the correlation between peptide sequence and tPAL events, we tested a set of peptides with periodic repeat units. For instance, EFEFEFC, composed of alternating glutamic acid and phenylalanine residues, generates an alternating pattern of low-noise and high-noise events during sequential cleavage (**Figure RL 2.13a**). Similarly, the reverse sequence FEFEFEC produced a signal pattern identical to EFEFEFC, but in

the opposite order (**Figure RL 2.14**). Additionally, for EFTREFTREFTRC, a peptide composed of repeating EFTR sequence units, we also observed a clear periodic signal pattern. Within each repeat unit, the peptide events exhibited similar fluctuation patterns when the same N-terminal residue was exposed, with only negligible variations. The only exception is the aa12 stage, where the signal differs significantly from aa4 and aa8 stages due to the differences in the downstream sequence, further confirming that the first few residues collectively play a dominant role in determining tPAL event features.

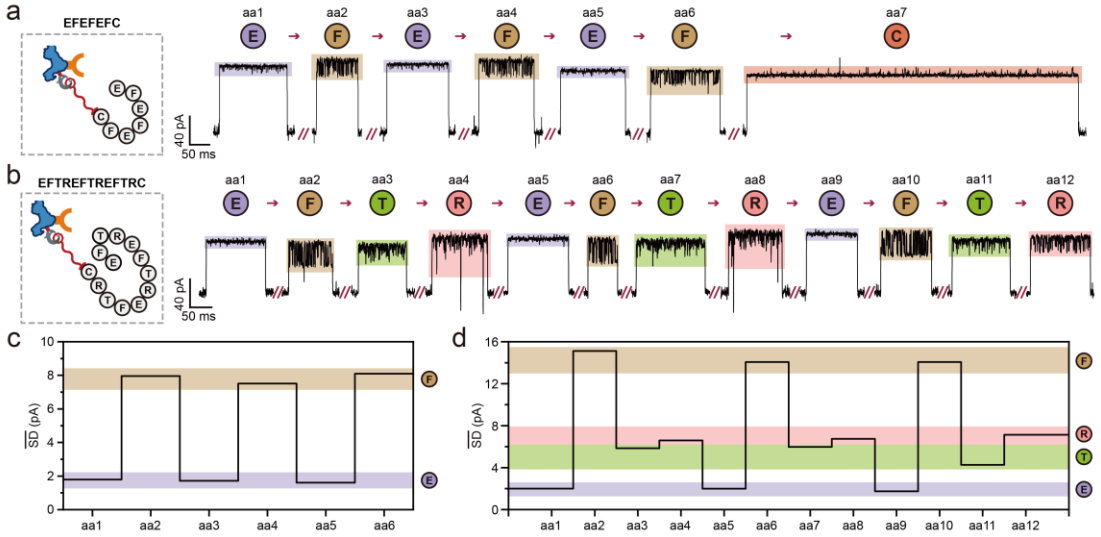


Figure RL 2.13. Sequential cleavage of peptides containing periodically repeating units. (**a, b**) Representative events obtained by sequential cleavage of (**a**) EFEFEFC and (**b**) EFTREFTREFTRC. All measurement was performed as described in Methods section. For EFTREFTREFTRC, thirteen stages were completely detected during the real-time digestion process. However, only the first twelve stages (aa1-aa12) are presented here due to space constraints. The corresponding N-terminal amino acid of each stage is marked above each trace. (**c**) The plot of $\overline{SD}$ at aa1-aa6 stages derived from EFEFEFC sequential cleavage. (**d**) The plot of $\overline{SD}$ at aa1-aa12 stages derived from EFTREFTREFTRC sequential cleavage.

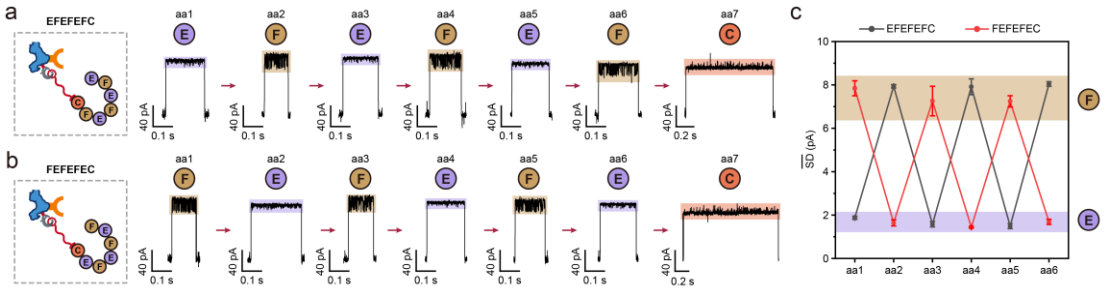


Figure RL 2.14. Sequential cleavage of EFEFEFC and FEFEFEC. (**a, b**) Representative events acquired during the sequential cleavage of (**a**) EFEFEFC and (**b**) FEFEFEC. The measurements were conducted as described in the **Methods** section using sgAP-dA15-chol. For each peptide, six enzymatic cleavages were observed, yielding seven distinct stages, which are respectively referred to as aa1-aa7. The corresponding N-terminal amino acid of each stage is labeled above the trace. (**c**) The plot of $\overline{SD}$ at aa1-aa6 stages derived from EFEFEFC and FEFEFEC sequential cleavage.

Drawing inspiration from the k-mer-based basecalling method in nanopore nucleic acid sequencing, we propose that a similar strategy can also be employed to determine amino acid sequences from tPAL data with the assistance of machine learning algorithm. To validate this, we constructed a miniaturized k-mer database using sequential cleavage events previously acquired from other peptides (**Figure RL 2.15**). A Wide Neural Network model was trained on this dataset, which achieved a validation accuracy of 95.9%.

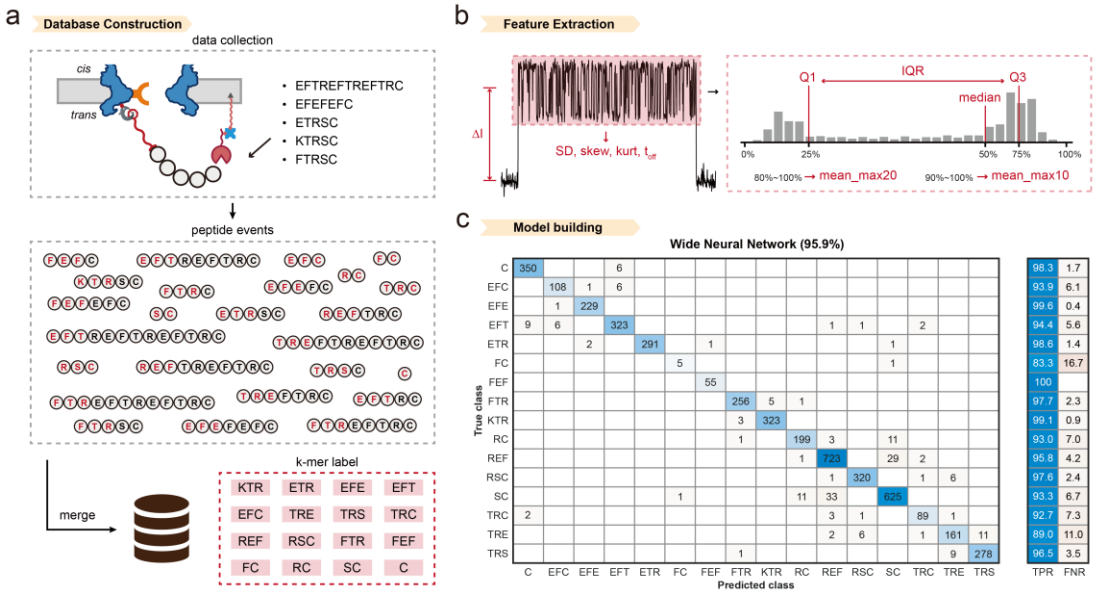


Figure RL 2.15. Construction of k-mer database and machine learning. (**a**) Database construction. Peptide events acquired during sequential cleavage of EFTREFTREFTRC, EFEFEFC, ETRSC, KTRSC and FTRSC were collected and merged to form the database. Peptide events with identical sequences in their first three residues are assigned the same 3-mer label, while events from dipeptides and amino acid are assigned corresponding 2-mer and 1-mer labels. (**b**) Feature extraction. Eleven event features were extracted from events to form a feature matrix, including ΔI, SD, $skew$, $kurt$, t_{off}, $median$, IQR, $Q1$, $Q3$, $mean\_max10$, $mean\_max20$. Here, $mean\_max10$ is defined as the mean value of top 10 percent of the current blockage ranked from largest to smallest, and $mean\_max20$ is defined as the mean value of top 20 percent of the current blockage ranked from largest to smallest. (**c**) The confusion matrix generated by the wide neural network model, which achieves a validation accuracy of 95.9%.

As a proof of concept, the data acquired during sequential cleavage of ETREFTRC was analyzed (**Figure RL 2.16b**). Here, all corresponding 3-mer components of ETREFTRC, such as ETR, TRE, REF and so on, can be extracted from results we previously obtained from other sequences, and the ETREFTRC sequence can be regarded as a sequence mixture thereof. The trained model was utilized to predict the k-mer identity of each stage, and the full peptide sequence could be reconstructed base on the overlap between adjacent k-mers (**Figure RL 2.16a**). To visualize the predicting results, we also calculated the amino acid probability for each position using the k-mer prediction results and generated an amino acid probability heatmap. As shown in **Figure RL 2.16c**, each amino acid residue in ETREFTRC was identified with high confidence.

Notably, the tPAL events at each stage contain information not only about the N-terminal

amino acid of the stage but also few downstream residues. Therefore, even if a small fraction of events at a certain stage is misclassified due to machine learning errors, the correct sequence can still be assembled through the contextual information of neighboring k-mers.

Specifically, we draw upon the approach of “de bruijn graph” used in genome assembly to reconstruct peptide sequences. As shown in **Figure RL 2.17**, even if some k-mers were incorrectly identified (marked with gray circles), they are unlikely to form complete connections with neighboring k-mers, and the correct sequence can still be reconstructed by selecting the optimal path that spans from the beginning to the end of the graph. This result further demonstrates the advantage of tPAL’s ability of peptide sequence decoding, which significantly improves recognition accuracy. However, direct peptide sequence decoding was not yet been done by any nanopore peptide sequencing approaches, to the best we know.

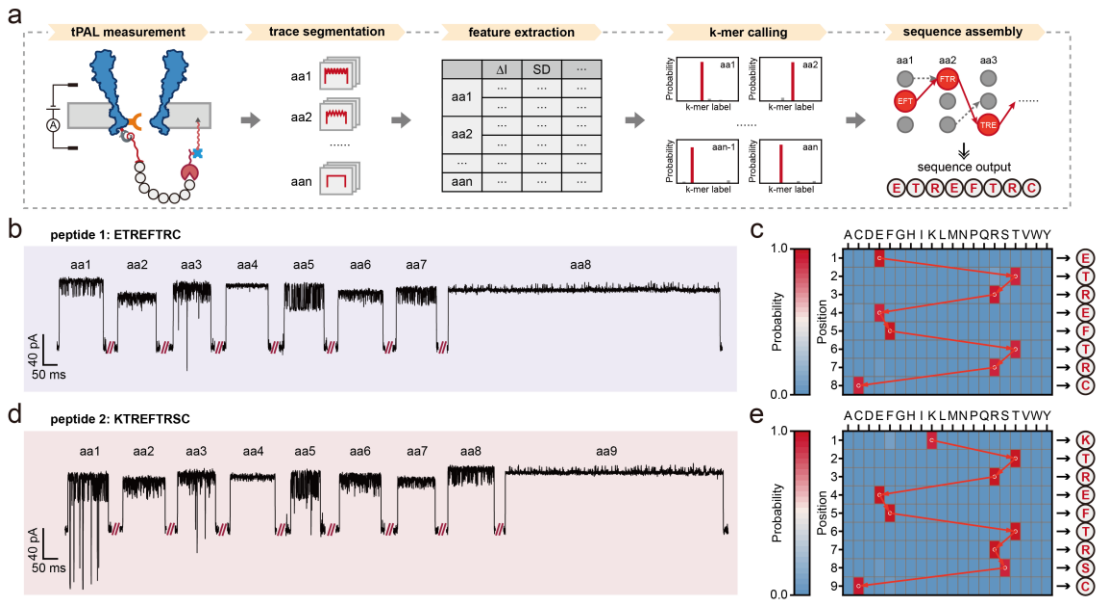


Figure RL 2.16. Peptide sequence decoding from tPAL events. (**a**) The workflow of machine learning-assisted sequence analysis. Briefly, the raw data obtained from peptide sequential cleavage is first divided into different stages. Feature extraction is then performed for sensing events within each stage to generate the feature matrix. Subsequently, a machine learning model pre-trained with a k-mer dataset is employed for event prediction, yielding k-mer labels and probability distributions for each stage. Finally, sequence assembly is conducted based on sequence overlap between k-mers from adjacent stages, outputting the complete peptide sequence. (**b, d**) Representative events acquired with (**b**) ETREFTRC and (**d**) KTREFTRSC during sequential cleavage. The measurements were performed as described in **Methods**. (**c, e**) Amino acid probability heatmaps derived from (**c**) ETREFTRC and (**e**) KTREFTRSC measurements. The probabilities of 20 amino acids at each position are calculated by normalizing the frequencies of all k-mers covering that position in prediction results. The optimal path (red polyline) connects the amino acid with the highest probability at each position, forming the predicted peptide sequence on the right side of each heatmap.

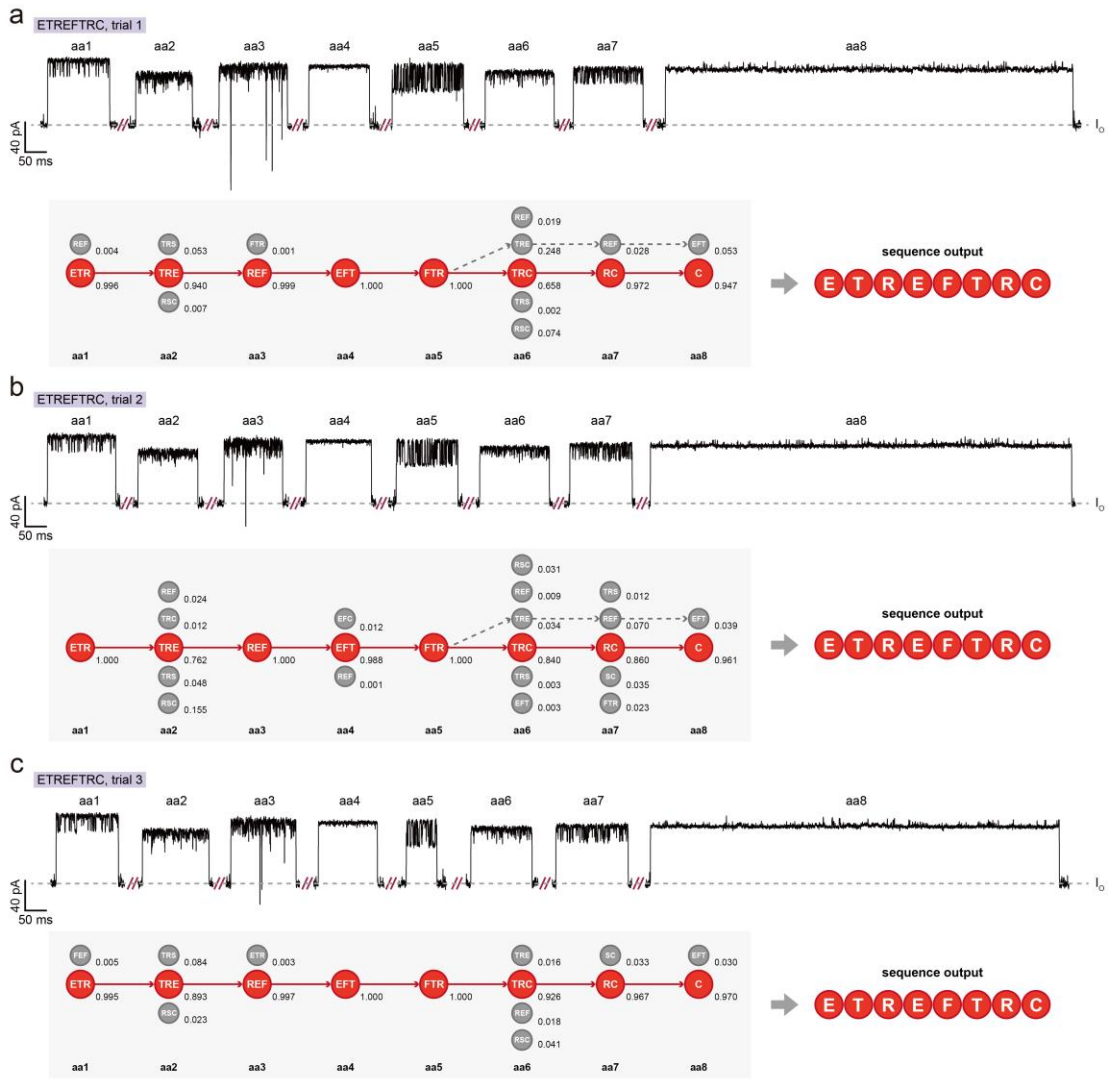


Figure RL 2.17. tPAL measurements of ETREFTRC and sequence assembly. Results were obtained from three independent experiments. All measurements were conducted according to the **Methods** section. Top: Representative traces acquired with ETREFTRC during sequential cleavage. Seven enzymatic cleavages were identified during the sequential cleavage of ETREFTRC, leading to eight distinct stages. Bottom: Sequence assembly graph. Each node in the graph represents a candidate k-mer predicted from the sensing events at corresponding stage, with their probability annotated alongside. Directed edges represent valid connections in which the suffix of a preceding k-mer exactly matches the prefix of the current k-mer. The optimal path, highlighted in red, is defined as the route exhibiting the highest cumulative probability while traversing exactly one node per stage. All remaining nodes and suboptimal paths are depicted in gray. The final peptide sequence was derived by sequentially merging the k-mers along this optimal path.

To verify the universality of this decoding method, we also analyzed the tPAL data of KTREFTRSC (**Figure RL 2.16d, e** and **Supplementary Figure 55**) and FTREFTRC (**Supplementary Figure 56**), and the correct sequences were successfully reconstructed in both cases.

In summary, in this round of revision, we have successfully demonstrated a data

processing pipeline for tPAL peptide sequencing and sequence decoding. Although tested in limited cases, the proof-of-concept for tPAL sequencing has been sufficiently demonstrated. We envision that with further optimization of the tPAL technology, expansion of the k-mer database, and improvement of machine learning algorithms, the tPAL method will eventually perfect the technique for later proteomics studies.

We would like to extend our sincere thanks once again to Reviewer 2 for your insightful comments, which have significantly enhanced the quality of our manuscript. In direct response to these valuable comments, we have incorporated additional materials and comprehensive discussions in the revised manuscript. For your convenience, the relevant section is pasted below. We remain open to further discussions from Reviewer 2 to help refine our methodology. The nanopore field eagerly awaits innovative solutions for peptide sequencing, and we value this opportunity to contribute these novel findings toward its ongoing development.

"Based on the insight that peptide events are predominantly influenced by the first few amino acids at the N-terminus, it is in principle possible to directly decode peptide sequence from tPAL results acquired during sequential enzymatic cleavage. … … Though demonstrated here using a miniaturized k-mer database constructed from events collected with existing peptides, further expansion of the k-mer database and optimization of machine learning algorithms hold considerable promise for practical sequencing applications."

Review Comments:

An additional point that could be explored with more experiments is the effect of peptide length. The authors tested only three peptides and did not address whether longer peptides—with potential structural features—would influence the signal, nor whether the peptidase can efficiently access longer substrates.

Response:

We deeply acknowledge Reviewer 2 for raising this critical comment regarding the influence of peptide length and structure. Notably, in the original manuscript, we have tested a total of 30 peptides instead of 3 peptides, with varying lengths and compositions, with the longest peptide consisting of 22 amino acid residues. Characteristic tPAL sensing events were still observed in this case (**Supplementary Figure 28**), indicating that tPAL method possesses the capability to detect long peptides with accessible N-terminus. In the revised manuscript, we further included more demonstrations and a total of 63 types of peptides were investigated in this single study.

Regarding your concern about the enzymatic cleavage efficiency of longer peptide, we have supplemented an experiment using a 13-mer peptide (EFTREFTREFTRC). The

result demonstrate that the aminopeptidase can efficiently access and progressively cleave this longer substrate, producing sequence-dependent signals (**Figure RL 2.13b**).

We also agree with reviewer's opinion that tPAL may struggle with highly structured peptides. If the N-terminus of a long peptide is buried within its structure, it cannot interact with the NTA-Ni adapter, thereby preventing signal generation. Moreover, peptidases may also struggle to access the N-terminus of folded peptides, which compromises their cleavage efficiency. We envision that future optimizations, such as employing denaturants or tuning the electroosmotic flow to unfold peptide structures, may extend this method to highly structured peptides.

Finally, we again acknowledge Reviewer 2 for raising this insightful discussion. To fully address your concern, we have incorporated extended discussions concerning the influence of peptide structure and potential solutions in the revised manuscript. For your convenience, we have included the relevant section below:

"For highly structured peptides which may fail to interact with NTA-Ni adapter and aminopeptidase, structural unfolding using electroosmotic flow EOF or denaturants can be employed to expose the N-terminus and enable detection."

Review Comments:

In summary, the results presented appear too preliminary to convincingly propose a method for sequencing peptides with nanopores.

Response:

We sincerely appreciate Reviewer 2 for all the valuable comments and suggestions. We fully acknowledge that our original submission presented an early iteration of the tPAL method, as our primary goal was to promptly share this significant breakthrough in single-molecule peptide sequencing with the academic community. However, following the constructive feedback from you and the other reviewers, we have substantially improved the methodology and thoroughly revised the manuscript. Key improvements include clarifying the superior resolution of our approach, demonstrating tPAL sequencing on long and repetitive sequences, and developing custom algorithms for direct sequence decoding. As the first of its kind, we never intended to report a commercial product-level demonstration of comprehensive nanopore peptide sequencing. However, we believe that these demonstrations provide a compelling and fundamentally novel way to truly sequence peptides at the single-molecule level. As part of this revision, we have added an **SI video clip** that presents **real-time playback of the core steps in peptide sequencing**, and we strongly encourage Reviewer 2 to watch this demonstration. We sincerely hope that our earnest revision efforts will earn us the opportunity to publish this work.

Notably, this unique technological configuration has effectively addressed unresolved

challenges such as the spatial resolution, peptide backsliding, sequence decoding, and the reliance on guide DNA or spacer peptides. Engineering efforts, such as the use of large-scale arrays, the construction of comprehensive databases, and the development of corresponding algorithms, can be pursued in later stages. Should the reviewer have further questions, we are more than willing to provide additional information or clarification. We are deeply grateful for your support and valuable feedback. We were pleased to hear from the editor that all four referees are experts deeply engaged in the nanopore field. We consider the nanopore community a united and collaborative group, and we are confident that the distinct advances reported herein will boost the long-term prosperity of our collective research endeavors. Once again, we thank you for all the insightful comments you have offered on our work. We would also be more than happy to address any further scientific inquiries regarding this manuscript at your convenience.

Reviewer #3:

Review Comments:

Protein sequencing has recently emerged as a central research focus within the nanopore community. Existing approaches can be broadly categorized into protease-assisted sequencing, fingerprinting, and strand sequencing, with strand sequencing generally considered the most promising for broad applications. Protease-assisted and fingerprinting methods, while valuable, often fall short of true sequencing due to capture preferences and limited generality across proteomes. In this context, Chen et al. present a strategy they term tPAL (transient Pore Analyte Looping). The authors engineer a hetero-octameric MspA nanopore with two bio-orthogonal reactive sites: one for NTA-Ni modification to capture the peptide N-terminus, and another for covalent tethering of the peptide near the pore constriction. Coupled with a cholesterol-modified aminopeptidase, this enables stepwise N-terminal trimming, multiple rereads of the same analyte, and discrimination of single amino acids, selected post-translational modifications (PTMs), and unnatural residues. The manuscript is clearly written. The technical achievement of constructing a dual-functionalized pore and demonstrating in-pore enzymatic stepping with single-residue resolution is commendable and will be of interest to the nanopore community. However, while these advances represent a step forward, I have reservations about the conceptual innovation, mechanistic clarity, and general applicability, which need to be addressed before the work can reach the bar for Nature. I recommend major revision.

Response:

We sincerely acknowledge Reviewer 3's enthusiastic evaluation and constructive comments on our manuscript. We are particularly grateful for the reviewer's recognition that this work represents a clear step forward and is of interest to the nanopore community, which is an extremely encouraging validation for a junior team like ours. We also appreciate the reviewer's acknowledgment of the technical achievements of our study. As the reviewer pointed out, tPAL is a novel sequencing strategy that fundamentally differs from existing methods. Unlike other strand-sequencing-based approaches, the tPAL strategy achieves strict single-amino-acid stepping, and the molecular adapter positioned at the pore constriction ensures high spatial resolution through specific chemical interactions with the peptide N-terminus. It also doesn't have any requirement of the charge of the target peptide. It also enables direct sequence interpretation as well, without any need of spacer peptide. These advantages highlight the innovative potential and practical value of tPAL in the field of peptide sequencing.

Regarding the reviewer's concerns about mechanistic clarity and general applicability, we sincerely acknowledge the reviewer for raising such rigorous suggestions. Many of these technical issues can be directly addressed through additional experiments. We sincerely apologize for overlooking these aspects in the original manuscript. In the revised version, we have also further included plenty of new results on nanopore sequencing and decoding,

further demonstrating how tPAL can be applied to the direct decoding of peptide sequences.

These technical improvements, largely inspired and guided by the reviewer's constructive feedback, have significantly strengthened the rigor and comprehensiveness of the manuscript. For the ease of your convenience, we have also provided detailed, point-by-point responses to your comments below.

Review Comments:

Enzyme dependence and generality. The method relies almost exclusively on sgAP, which shows sequence preferences and stalls at certain motifs (e.g., Pro or phosphorylated residues). This constrains its utility for proteome-wide sequencing. Demonstrating activity with an enzyme panel or cocktails would substantially strengthen the generality of the approach.

Response:

We sincerely appreciate Reviewer 3's constructive suggestion regarding the enzyme dependence. The reviewer is correct in pointing out that the current demonstration relies solely on sgAP. However, during the development of this technique, we actually have screened a wide range of aminopeptidases and found that sgAP performs the best. Notably, all tPAL measurements were conducted in a 1.5 M KCl buffer at room temperature, and we feel fortunate that sgAP retains its aminopeptidase activity under such high-salt conditions. Another technical advantage of sgAP is its relatively broad substrate suitability for amino acid cleavage, although, as the reviewer accurately noted, certain amino acids such as Xaa-Pro (Pro as the 2nd N-terminal amino acid) and phosphorylated residues are not cleavable or could not be efficiently cleaved.

We also feel fortunate that Reviewer 3 has made such an insightful suggestion regarding the use of an enzyme panel or cocktail, which is a brilliant idea for further broadening the generality of tPAL. Indeed, given the unique configuration of tPAL, it is technically compatible with such an enzyme panel strategy. Therefore, in the revised manuscript, we have included further demonstration of this approach, and we are extremely grateful for the reviewer's valuable suggestion.

As previously described, during sequential cleavage of phosphopeptide, when the phosphorylated residue is exposed at the N-terminus, characteristic events with low noise and high current level could be observed, allowing for the determination of the presence and position of peptide phosphorylation. However, due to the inability of sgAP to cleave phosphorylated residue, it fails to recognize the sequence downstream of the phosphorylation site. Following reviewer's suggestion, we designed a novel enzymatic strategy that combines the use of sgAP with phosphatase for real-time sequencing of phosphopeptide (**Figure RL 3.1a**).

As a proof-of-concept demonstration, we tested a peptide containing a phosphorylated tyrosine: L(pY)FTRSC (**Figure RL 3.1b**). Upon addition of sgAP, the N-terminal leucine residue was rapidly cleaved. Characteristic low-noise events were observed, and no further cleavage occurred, indicating the presence of the phosphorylation. Subsequently, lambda phosphatase was added. When the phosphate group was hydrolyzed, a distinct change of sensing events was clearly observed. The characteristics of these events were consistent with those observed at aa2 stage during the sequential cleavage of non-phosphopeptide (LYFTRSC), further confirming that the signal change resulted from the removal of phosphate group. After that, five consecutive cleavages were detected, completely hydrolyzing the peptide to a single cysteine residue eventually.

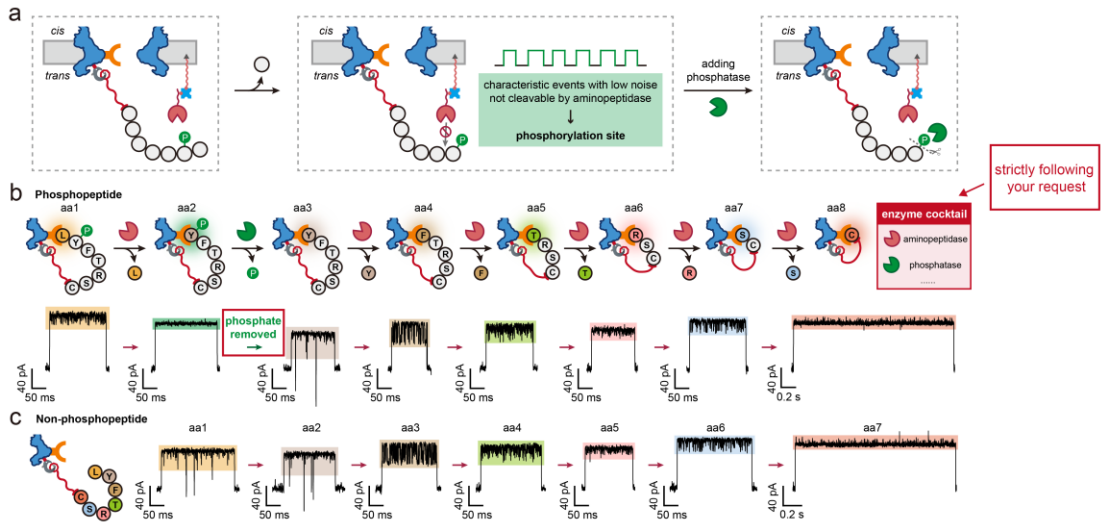


Figure RL 3.1. Sequential cleavage of phosphopeptide using enzyme cocktail. (**a**) Schematic diagram depicting sequential cleavage of phosphopeptide using both aminopeptidase and phosphatase. Initially, sgAP-dA15-chol was employed to sequentially cleave the immobilized peptide. As digestion progresses to the phosphorylation site, characteristic low-noise events were detected. Due to the resistant of phosphorylated amino acid residues to aminopeptidase, a cleavage inhibition was observed at this stage. Based on these observations, it can be inferred that the N-terminal amino acid at this point bears a phosphorylation modification. Phosphatase was therefore introduced to remove the phosphate group, enabling the aminopeptidase to proceed further. (**b**) The mechanism (top) and corresponding events (bottom) acquired during the sequential cleavage of L(pY)FTRSC. sgAP-dA15-chol (8 nM) and its cofactors (5 μM Zn^{2+} and 1 mM Ca^{2+}) were first added to the *trans* chamber. Upon detection of the phosphorylation modification at the N-terminus, lambda phosphatase (1.4 kU), along with its cofactor (1 mM Mn^{2+}) was added to dephosphorylate the N-terminal residue, allowing sgAP to keep cleaving the immobilized peptide. (**c**) Representative events acquired during the sequential cleavage of LYFTRSC. With the assistant of sgAP-dA15-chol, seven distinct stages were observed.

We believe that similar approaches could be applied to other sequence motifs that are resistant to sgAP cleavage. However, the candidate enzymes must also exhibit salt resistance and remain active at room temperature. In future work, we will continue to explore additional enzyme combinations and incorporate AI-assisted enzyme engineering techniques to more effectively address the issue of enzymatic selectivity. It is also feasible to fuse phosphatase or glycosidase with sgAP to enable more efficient cleavage of specific

post-translational modifications in specialized sequencing applications. To fully address this issue, we have added results as suggested by the reviewer along with relevant discussions in the revised manuscript. For the ease of reviewer's reference, the added discussions are also pasted below:

*"…An enzyme cocktail strategy was applied for phosphopeptide measurement (**Supplementary Figure 50a**). Briefly, aminopeptidase sgAP was firstly employed to sequentially cleave the anchored peptide L(pY)FTRSC, identifying amino acid residues upstream of the phosphorylation site. The observation of characteristic low-noise events, coupled with the absence of further cleavages, indicates that the phosphorylated residue has reached the N-terminus of the peptide. At this point, lambda phosphatase was introduced to specifically hydrolyze the phosphate group at the N-terminal residue, thereby enabling sgAP to continue cleaving the peptide to achieve full length tPAL measurement (**Supplementary Figure 50b, c**)."*

Review Comments:

Error rates and blinded decoding. Reported high accuracies are derived from curated datasets. No blinded decoding of mixed peptides, no confusion matrices, and no tests under varied experimental conditions are presented. These experiments are essential to evaluate real-world sequencing performance.

Response:

We sincerely appreciate Reviewer 3's insightful comments regarding error rate and blinded decoding, which are crucial for the practical application of our method in real-world. In tPAL measurement, the peptide can be read at single amino acid step through sequential digestion by aminopeptidase, and the nanopore events at each stage is strongly associated with its sequence, especially the first a few residues at the N-terminus. In this round of revision, drawing inspiration from the k-mer-based basecalling method in nanopore nucleic acid sequencing, we propose that a similar strategy can also be employed to determine amino acid sequences from tPAL data with the assistance of machine learning algorithm.

To validate this, we constructed a miniaturized k-mer database using sequential cleavage events from several peptides (**Figure RL 3.2**). A wide neural network model was trained on this dataset, with a validation accuracy of 95.9%. As a proof of concept, the data acquired from ETREFTRC sequential cleavage was analyzed (**Figure RL 3.3b**). Since tPAL can only measure a single peptide at a time, we designed this demonstration using a peptide of a specific sequence, which is **a mix of k-mer sequences** among the training data.

The trained model was utilized to predict the k-mer identity of each stage, and the full peptide sequence could be reconstructed base on the overlap between adjacent k-mers

(**Figure RL 3.3a**). To visualize the predicting results, we also calculated the amino acid probability for each residue position using the k-mer prediction results and generated an amino acid probability heatmap. As shown in **Figure RL 3.3c**, each amino acid residue in ETREFTRC was identified with extremely high confidence.

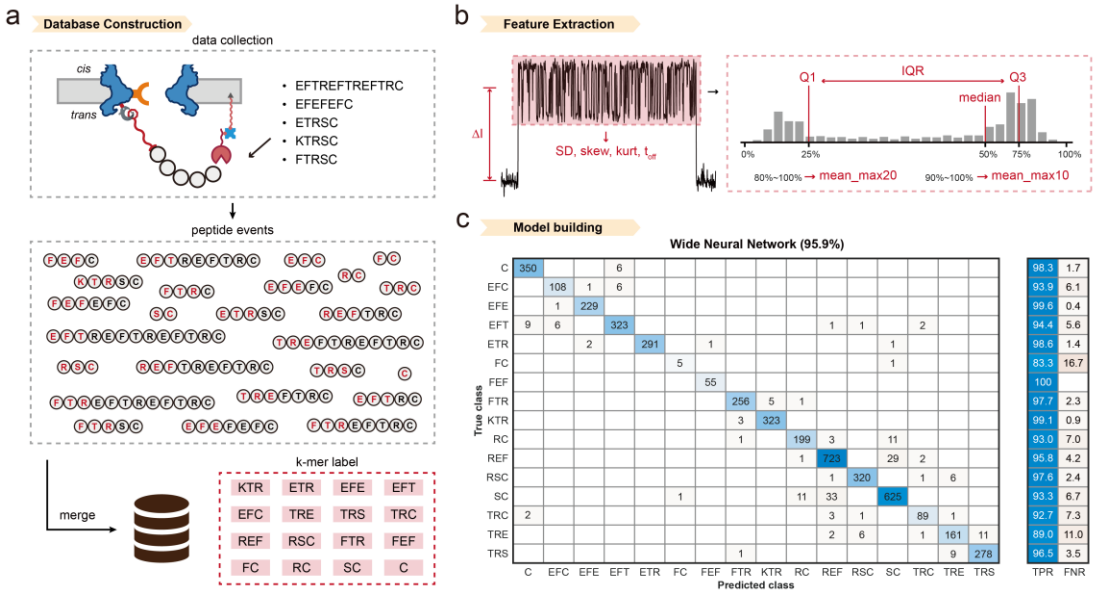


Figure RL 3.2. Construction of k-mer database and machine learning. (**a**) Database construction. Peptide events acquired during sequential cleavage of EFTREFTREFTRC, EFEFEFC, ETRSC, KTRSC and FTRSC were collected and merged to form the database. Peptide events with identical sequences in their first three residues are assigned the same 3-mer label, while events from dipeptides and amino acid are assigned corresponding 2-mer and 1-mer labels. (**b**) Feature extraction. Eleven event features were extracted from events to form a feature matrix, including ΔI, SD, $skew$, $kurt$, t_{off}, $median$, IQR, $Q1$, $Q3$, $mean\_max10$, $mean\_max20$. Here, $mean\_max10$ is defined as the mean value of top 10 percent of the current blockage ranked from largest to smallest, and $mean\_max20$ is defined as the mean value of top 20 percent of the current blockage ranked from largest to smallest. (**c**) The confusion matrix generated by the wide neural network model, which achieves a validation accuracy of 95.9%.

Notably, the tPAL events at each stage contain information not only about the N-terminal amino acid of that stage but also downstream residues. Therefore, even if a small fraction of events at a certain stage is misclassified due to machine learning errors, the correct sequence can still be assembled through the contextual information of neighboring k-mers. Specifically, we draw upon the approach of de bruijn graph used in genome assembly to reconstruct peptide sequences. As shown in **Figure RL 3.4**, even if some k-mers were incorrectly identified (marked with gray circles), they are unlikely to form complete connections with neighboring k-mers, and the correct sequence can still be reconstructed by selecting the optimal path that spans from the beginning to the end of the graph. This result further demonstrates the advantage of tPAL's ability of peptide reread, which significantly improves recognition accuracy.

To verify the universality of this decoding method, we also analyzed the tPAL data of KTREFTRSC (**Figure RL 3.3d, e** and **Supplementary Figure 55**) and FTREFTRC

(**Supplementary Figure 56**), and the correct sequences were successfully reconstructed in both cases. tPAL can also in principle deal with a mixture of peptides. However, due to its measurement nature, only one peptide can be measured a time. However, the demonstrations with ETREFTRC, FTREFTRC and KTREFTRSC have approved its discrimination power when variable sequences of peptides were fed to the sensor, demonstrating a variable experiment condition.

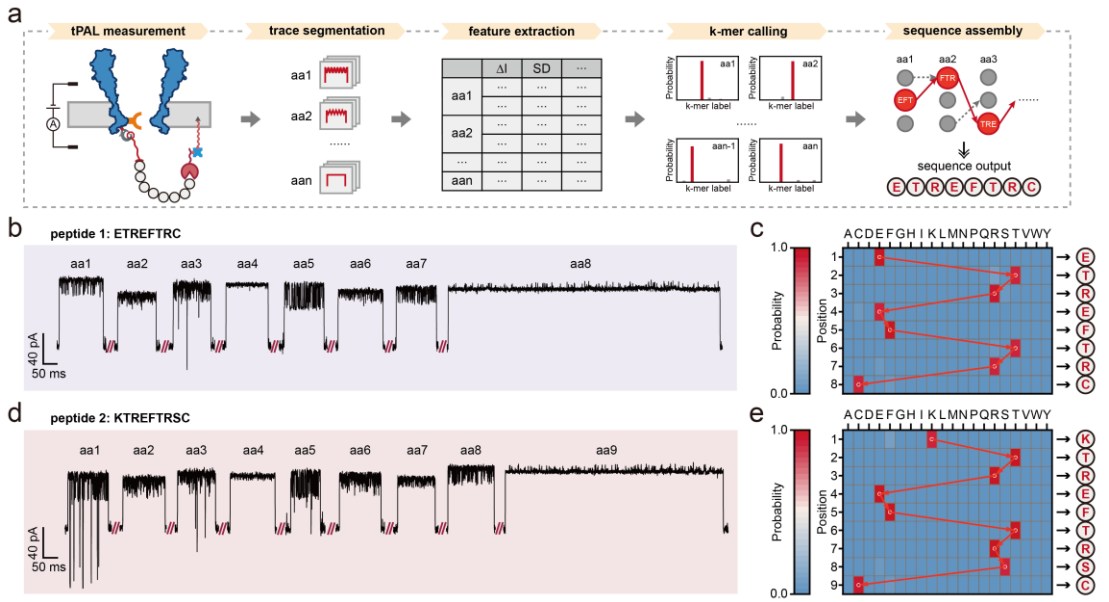


Figure RL 3.3. Peptide sequence decoding from tPAL events. (**a**) The workflow of machine learning-assisted sequence analysis. Briefly, the raw data obtained from peptide sequential cleavage is first divided into different stages. Feature extraction is then performed for sensing events within each stage to generate the feature matrix. Subsequently, a machine learning model pre-trained with a k-mer dataset is employed for event prediction, yielding k-mer labels and probability distributions for each stage. Finally, sequence assembly is conducted based on sequence overlap between k-mers from adjacent stages, outputting the complete peptide sequence. (**b, d**) Representative events acquired with (**b**) ETREFTRC and (**d**) KTREFTRSC during sequential cleavage. The measurements were performed as described in **Methods**. (**c, e**) Amino acid probability heatmaps derived from (**c**) ETREFTRC and (**e**) KTREFTRSC measurements. The probabilities of 20 amino acids at each position are calculated by normalizing the frequencies of all k-mers covering that position in prediction results. The optimal path (red polyline) connects the amino acid with the highest probability at each position, forming the predicted peptide sequence on the right side of each heatmap.

In summary, we have successfully developed a specialized data processing pipeline for tPAL method. Although tested in limited cases, the proof of concept for peptide decoding has been sufficiently demonstrated, including quantitative evaluation with confusion matrix and error rates. We envision that with further optimization of the tPAL technology, expansion of the k-mer database, and improvement of machine learning algorithms, the tPAL method will be applicable to real-world sequencing scenarios.

We thank reviewer 3 again for raising this insightful discussion of peptide decoding. Relevant results and discussions have been added to the revised manuscript to fully address this aspect. For the ease of your reference, the relevant content is included below:

"Based on the insight that peptide events are predominantly influenced by the first few amino acids at the N-terminus, it is in principle possible to directly decode peptide sequence from tPAL results acquired during sequential enzymatic cleavage. … … Though demonstrated here using a miniaturized k-mer database constructed from events collected with existing peptides, further expansion of the k-mer database and optimization of machine learning algorithms hold considerable promise for practical sequencing applications."

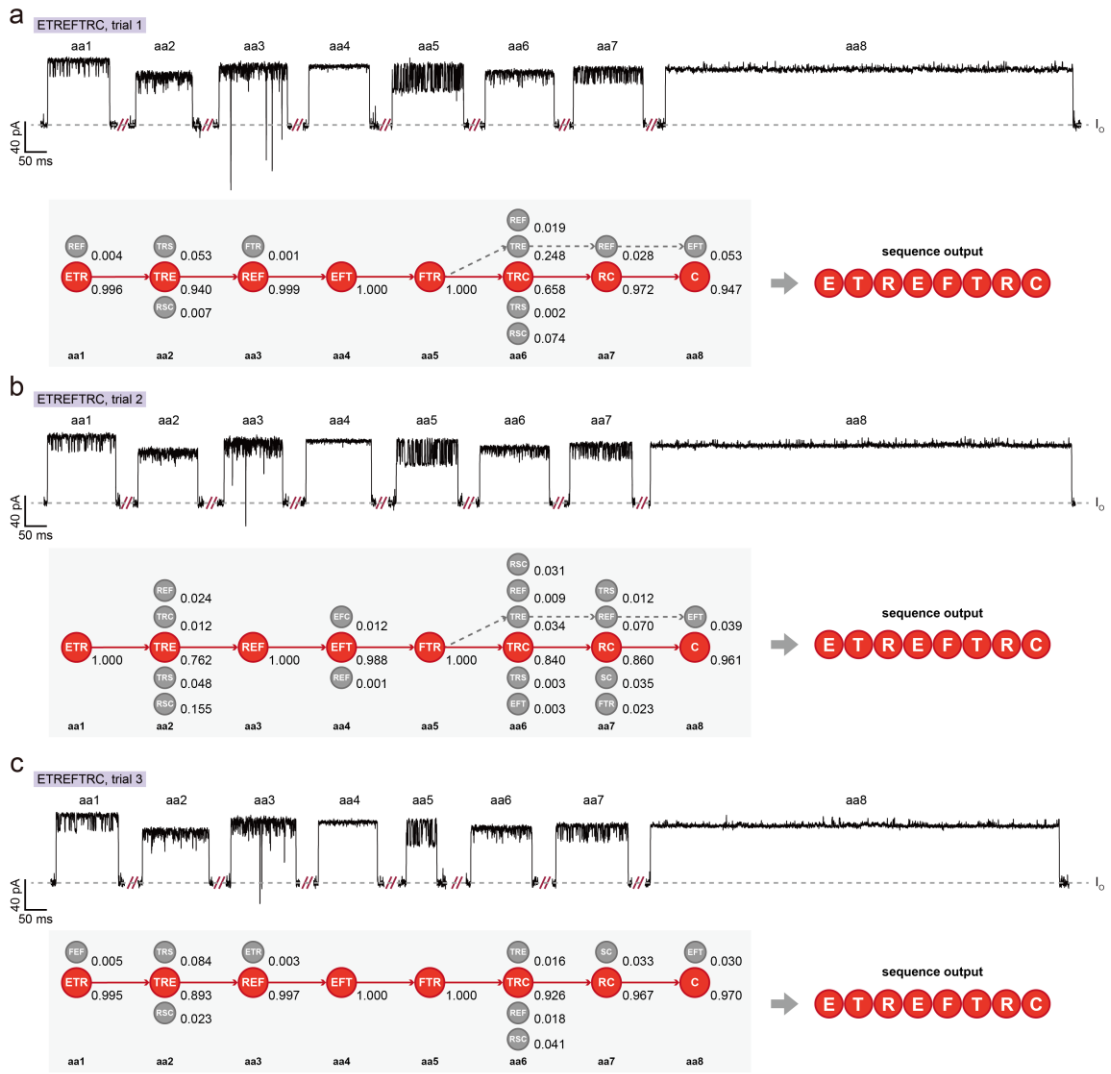


Figure RL 3.4. tPAL measurements of ETREFTRC and sequence assembly. Results were obtained from three independent experiments. All measurements were conducted according to the **Methods** section. Top: Representative traces acquired with ETREFTRC during sequential cleavage. Seven enzymatic cleavages were identified during the sequential cleavage of ETREFTRC, leading to eight distinct stages. Bottom: Sequence assembly graph. Each node in the graph represents a candidate k-mer predicted from the sensing events at corresponding stage, with their probability annotated alongside. Directed edges represent valid connections in which the suffix of a preceding k-mer exactly matches the prefix of the current k-mer. The optimal path, highlighted in red, is defined as the route exhibiting the highest cumulative probability while traversing exactly one node per stage. All remaining nodes and suboptimal paths are depicted in gray. The final peptide sequence was derived by sequentially merging the k-mers along this optimal path.

Review Comments:

PTM coverage. The study demonstrates detection of a small number of PTMs, but does not assess isobaric or chemically similar modifications (e.g., Lys acetylation vs. methylation). Expanded PTM analysis is needed to support broader claims.

Response:

We deeply appreciate Reviewer 3's constructive suggestion regarding the expansion of PTM analysis. In response to your request, we have conducted additional experiments to evaluate the detection of lysine trimethylation (Kme3), a PTM with isobaric mass to lysine acetylation (Kac). Specifically, we tested a dodecapeptide named Pep12Kme3 (L(Kme3)LRGDDGSESC), and compared it with Pep12Kac (L(Kac)LRGDDGSESC). While lysine trimethylation and acetylation share nearly identical molecular weights, making them challenging to be distinguished using conventional mass spectrometry, this nanopore strategy however has sufficient resolution to fully discriminate between these two PTMs with no ambiguity (**Figure RL 3.5**), further demonstrating the high resolution of tPAL identifying chemically similar PTMs and suggesting its potential for the discovering of previously un-discovered PTMs.

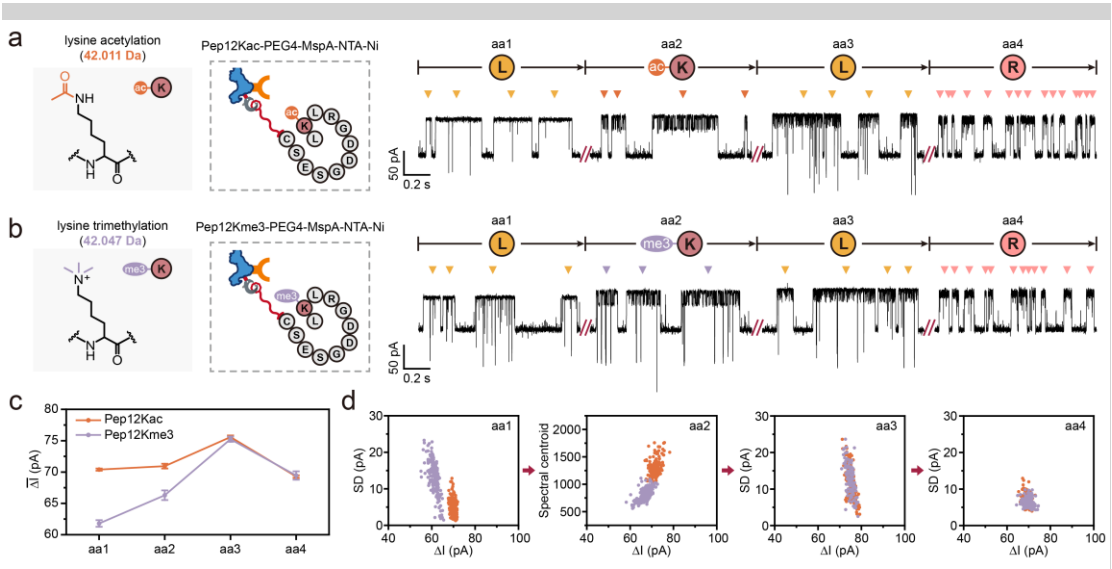


Figure RL 3.5. Discrimination of peptides containing isobaric PTMs. (**a**) The chemical structure of lysine acetylation and representative traces acquired with Pep12Kac sequential cleavage. (**b**) The chemical structure of lysine trimethylation and representative traces acquired with Pep12Kme3 sequential cleavage. The measurements were performed as described in **Methods**. (**c**) Comparison of $\overline{\Delta I}$ at each stage between Pep12Kac and Pep12Kme3. Data are presented as mean ± standard deviation values derived from three independent measurements. (**d**) The scatter plots of results respectively acquired with Pep12Kac and Pep12Kme3.

Following your kind suggestion, we have well demonstrated the recognition of methionine oxidation, tyrosine phosphorylation, lysine acetylation, and lysine trimethylation using tPAL, all in this single paper.

Although not explicitly demonstrated in the revised manuscript, this method is also capable of detecting other PTMs, such as sulfation and glycosylation. However, due to the enzymatic selectivity of the aminopeptidase used in current system, we are temporarily unable to perform stepwise enzymatic cleavage through these PTM sites. However, in future prospects, aminopeptidase engineering may address this limitation.

We again sincerely acknowledge this valuable inquiry from Reviewer 3. To fully address this issue, the newly acquired results along with relevant discussions of PTM differentiation were added to the revised manuscript. The added discussion is also pasted below for the ease of your reference:

"To demonstrate PTM identification using tPAL, we tested Pep12Mo (L(Mo)LRGDDGSESC), Pep12Kac (L(Kac)LRGDDGSESC) and Pep12Kme3 (L(Kme3)LRGDDGSESC), which contain oxidation, acetylation and trimethylation modifications, respectively"

*"Notably, lysine trimethylation and acetylation, which are a pair of PTMs with nearly identical mass, can be clearly differentiated through tPAL method (**Supplementary Figure 44**)."*

Review Comments:

Sensing mechanism. Several aspects of the physical basis for current signatures remain unclear: a) Which forces (electro-osmotic flow, electrophoretic force, or both) govern loop configuration inside the pore?

Response:

We sincerely appreciate Reviewer 3 for raising this important discussion regarding the physical basis of our sensing mechanism. In the tPAL configuration, the C-terminus of the analyte peptide is anchored near the pore constriction, resulting in an extremely close spatial proximity between the peptide N-terminus and the NTA-Ni adapter. The chemical reaction between the peptide N-terminus and the NTA-Ni adapter would spontaneously and reversibly happen, forming the basis of tPAL sensing. This spatial proximity increases the probability of intermolecular collisions, thereby significantly enhances the probability of loop formation.

To experimentally demonstrate this, tPAL sensing was conducted at an extremely low voltage (+5 mV). An electrically neutral peptide, LMC, was selected as the model analyte, and spontaneous appearance of corresponding events were still clearly observed under this minimal voltage (**Figure RL 3.6**), further confirming that spatial proximity promotes loop formation. However, with such low applied voltage and the electrically neutral peptide, the external force acting on the peptide should be already negligible. Though the generation of tPAL events is spontaneous, other forces, such as electrophoresis and

electroosmotic flow may also act to modulate the event appearance rate. A core advantage of this tPAL sensing mechanism is that the same target peptide could be measured repeatedly at different voltage conditions to produce richer sensing information (**Extended Data Figure 1**), a sensing mode we refer to as multi-voltage sweeping. To fully address this issue, a newly added SI Figure and relevant discussions have been incorporated into the revised manuscript and are also provided below for your reference:

*"Even at an extremely low applied voltage of +5 mV, tPAL events were still observable (**Supplementary Figure 10**), indicating that their generation is governed by spontaneous chemical reactions. Electrophoretic forces and electroosmotic flow, however, may also act to regulate the event appearance frequency."*

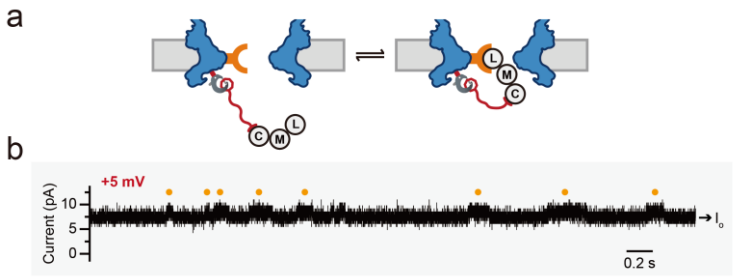


Figure RL 3.6. tPAL sensing at +5 mV. (**a**) The sensing mechanism of tPAL. (**b**) A representative trace of LMC-PEG4-MspA-NTA-Ni. The measurement was performed in a buffer of 1.5 M KCl, 10 mM CHES, pH 9.0 with a continually applied voltage of +5 mV. The yellow circles annotate tPAL events.

Review Comments:

How do signals compare between tethered and untethered peptides, given their different conformational freedoms?

Response:

We sincerely appreciate Reviewer 3's insightful comment, which has raised valuable discussions regarding the sensing mechanism of our method. To fully address your concern, we have tested more peptides (LRGESC, RGESC, GESC, ESC, SC and C) in the untethered state (**Figure RL 3.7**). These results indicate that, for long peptides, the tPAL sensing events exhibit highly similar noise feature to those in stochastic sensing, despite differences in their event amplitudes. However, as the peptide length decreases, the differences between stochastic sensing events and tPAL sensing events become more significant.

Actually, the tPAL strategy employs a flexible PEG linker to connect the peptide with the nanopore. Therefore, we propose that the C-terminal tethering of peptide imposes minimal constraints on its N-terminal region, allowing it to retain considerable conformational freedom and interact freely with the NTA-Ni adapter. For longer peptides, the tPAL sensing events is predominantly determined by the first few N-terminal amino acids, whereas the

contribution from the C-terminally attached PEG linker is relatively minor. Consequently, the sensing events of longer peptides in their untethered state and tethered state exhibit notable similarity. However, when the peptide is shortened, the C-terminal PEG linker becomes closer to the recognition site, thereby exerting a more pronounced influence on the characteristics of the sensing events. This contributes to the observed signal discrepancy in short peptides and amino acid, such as SC and C.

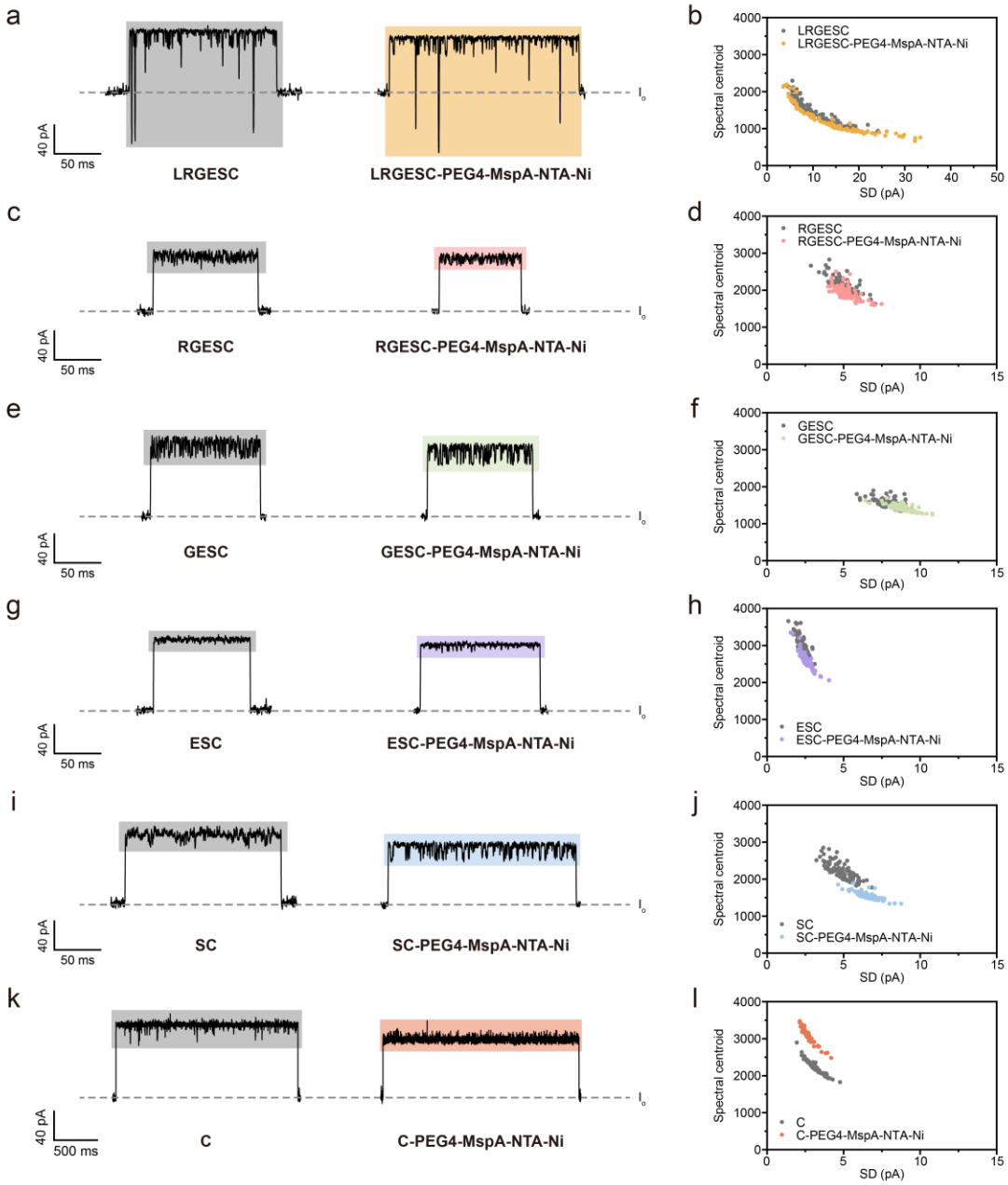


Figure RL 3.7. Comparation of events acquired by tPAL sensing and stochastic sensing. (**a, c, e, g, i, k**) Left: Representative events of (**a**) LRGESC, (**c**) RGESC, (**e**) GESC, (**g**) ESC, (**i**) SC and (**k**) C measured by (N90NTA-Ni, N93AzK)$_1$(M2)$_7$. Right: Representative events acquired with (**a**) LRGESC-PEG4-MspA-NTA-Ni, (**c**) RGESC-PEG4-MspA-NTA-Ni, (**e**) GESC-PEG4-MspA-NTA-Ni, (**g**) ESC-PEG4-MspA-NTA-Ni, (**i**) SC-PEG4-MspA-NTA-Ni and (**k**) C-PEG4-MspA-NTA-Ni. All measurements were performed in a buffer of 1.5 M KCl, 10 mM CHES, pH 9.0 with a continually applied voltage of +100 mV. (**b, d, f, h, j, l**) Event scatter plots of ΔI versus $spectral\ centroid$ for events

acquired by stochastic sensing of (**b**) LRGESC, (**d**) RGESC, (**f**) GESC, (**h**) ESC, (**j**) SC and (**l**) C and their corresponding tPAL sensing results. For each condition, 100 events were used to generate the scatter plots.

To fully address this issue, these newly acquired results, as kindly suggested by the reviewer, along with relevant discussions have been added to the revised manuscript. For the ease of your reference, these discussions are also pasted below:

*"The fluctuation characteristics of these binding events acquired by tPAL closely resemble those observed with stochastic sensing (**Supplementary Figure 14**), suggesting that the C-terminal PEG4 linker has minimal impact on the signal of these peptides and the event features are predominantly influenced by a few amino acids at the N-terminus."*

"This differentiation should arise from the increased contribution of the C-terminal PEG linker to the signal as the peptide shortens."

Review Comments:

Although the authors note multiple residues may influence the current, Figure 5 suggests signals may often be dominated by the N-terminal residue. Clarification of how many residues collectively define a signal is crucial.

Response:

We sincerely appreciate reviewer 3's insightful and constructive comment regarding the spatial resolution of tPAL. We are grateful for the opportunity to clarify this in more detail.

Experimentally, we have conducted a systematic investigation using three peptides, including LEFTNSC, LEFTLSC and LEFTFSC, which differ in sequence only at the fifth position. During sequential cleavage, seven distinct event stages were detected for each peptide, denoted as aa1 to aa7 (**Figure RL 3.8**). Comparative analysis of the signals generated during the cleavage revealed significant differences in these peptides at aa3-aa5 stages, which were directly attributed to the single amino acid mutation at fifth position. However, the signals in aa1 and aa2 stages are almost identical across all three peptides, indicating that events in the initial two stages are largely independent of the mutation.

To further validate this conclusion, we also conducted additional comparative analysis of other peptides. As shown in **Figure RL 3.9**, peptides sharing identical N-terminal sequences in the first three residues exhibit highly similar event patterns. However, subtle differences in signals were also observed between corresponding peptides. These results indicate that the first three N-terminal residues are the primary contributors to peptide events. While subsequent residues may also play a minor role, their influence diminishes as the position moves away from the N-terminus.

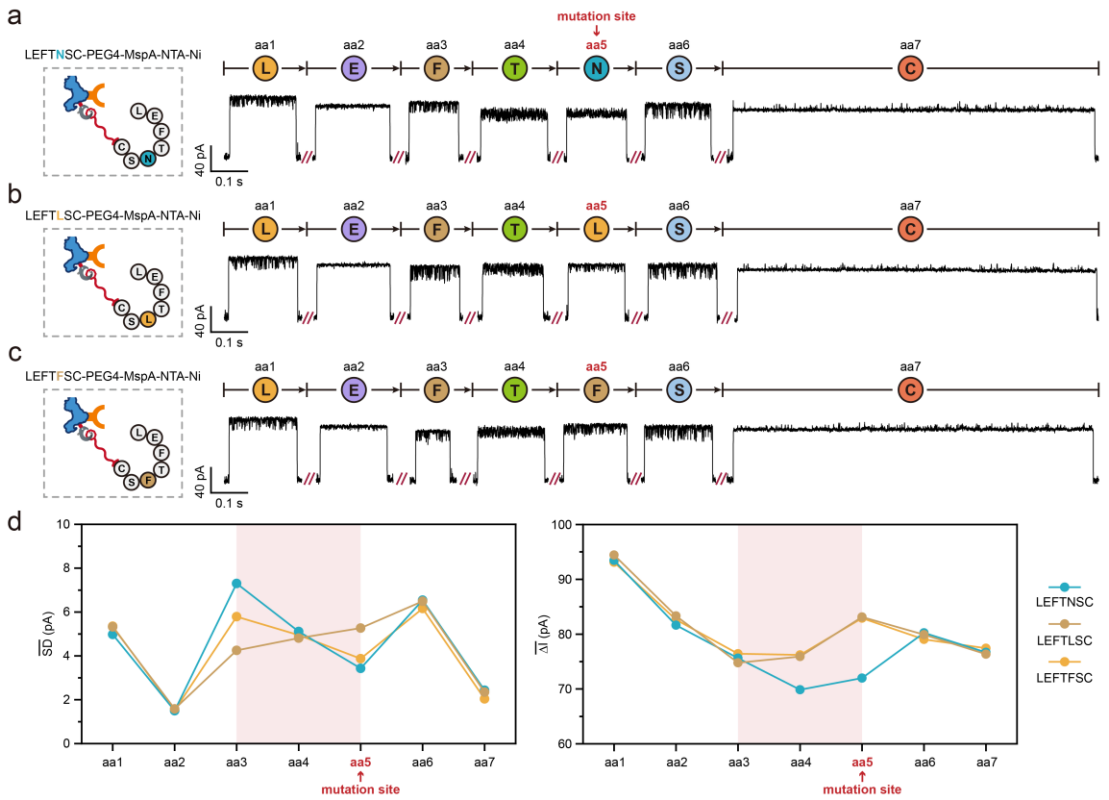


Figure RL 3.8. Sequential cleavage of peptides with single amino acid variation at the fifth position. (**a-c**) Representative events obtained by sequential cleavage of (**a**) LEFTNSC, (**b**) LEFTLSC and (**c**) LEFTFSC. The measurements were performed as described in Methods. For each peptide, six steps of enzymatic cleavages were observed, generating seven stages referred to as aa1-aa7, respectively. (**d**) Comparison of $\overline{SD}$ (left) and $\overline{\Delta I}$ (right) at different stages between LEFTNSC, LEFTLSC and LEFTFSC.

In response to reviewer's constructive feedback, we have thoroughly revised the manuscript to include these findings and further clarify the spatial resolution of tPAL method. The updated discussions are pasted below for your reference:

*"To evaluate the spatial resolution of tPAL, three peptides that differ only at the fifth residue, including LEFTNSC, LEFTLSC, and LEFTFSC, were tested (**Supplementary Figure 52**). A comparison of the events generated during enzymatic cleavage revealed distinct features among the three peptides at the aa3-aa5 stages, which can be directly attributed to the mutation at the fifth position. However, the signals at the aa1 and aa2 stages were highly similar across all three peptides, indicating that the events at these stages are not affected by the mutation. Additional cases with identical N-terminal amino acid compositions for the first three residues also showed high signal similarity (**Supplementary Figure 53**), suggesting that the first three N-terminal amino acids play a dominant role in determining event features, with only minor contributions from subsequent residues. Moreover, the influence of a given residue decreases progressively with increasing distance from the N-terminus."*

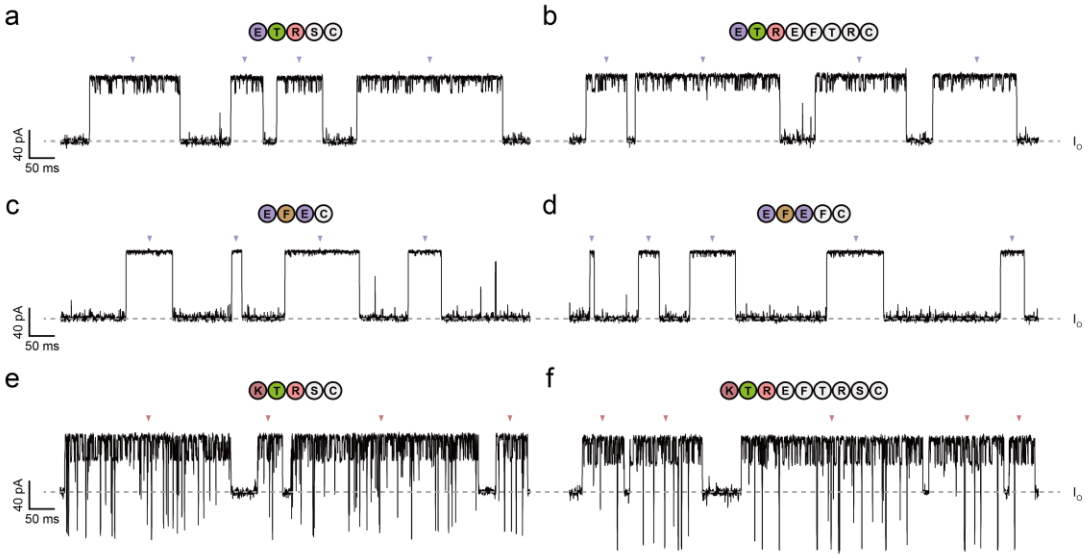


Figure RL 3.9. Comparison of tPAL events of peptides with identical first three residues. (**a, b**) Comparison of tPAL events of (**a**) ETRSC and (**b**) ETREFTRC. (**c, d**) Comparison of tPAL events of (**c**) EFEC and (**d**) EFEFC. (**e, f**) Comparison of tPAL events of (**e**) KTRSC and (**f**) KTREFTRSC.

Review Comments:

Can tethered peptides enter the pore without NTA-Ni^{2+} coordination, and if so, how often? The observed transient spikes need to be explained in this context.

Response:

We sincerely appreciate reviewer 3 for raising this inspiring discussion. We sincerely apology for overlooking this issue simply because only the coordination between the NTA-Ni^{2+} and the peptide could produce sequence dependent sensing information. To fully address your request, we performed tPAL measurements under conditions where only NTA modification was applied without Ni^{2+} modification (**Figure RL 3.10a**). In this case, the tethered peptide cannot form coordination interactions when entering the pore (**Figure RL 3.10b**). At -100 mV, long-term blocking events from the negatively charged peptides LGEDSC and Pep22 can be clearly observed (**Figure RL 3.10c-e**). Under +100 mV, no obvious signal was observed for LGEDSC-PEG4-MspA-NTA (**Figure RL 3.10e**), while Pep22-PEG4-MspA-NTA produced many negative-going spikes (**Figure RL 3.10d**). These spike events were also observed in tPAL measurement of Pep22 with NTA-Ni adapter (**Supplementary Figure 28c**), which may be attributed to the transient blockage of long peptide.

Based on above results, we concluded that the tethered peptide can enter the pore without coordination interaction. It however only generates extremely short-residing spiky events and won't interfere with the actual tPAL measurement. We sincerely apologize for ignoring this discussion in the original manuscript. Following your helpful suggestion, we have incorporated additional materials and extended discussions in the revised version.

*"Notably, in the case of Pep22, downward spikes were observed in addition to characteristic tPAL events. These spike events are attributed to transient translocation of the tethered peptide. Measurements conducted without Ni^{2+} modification further confirmed the non-coordination mechanism (**Supplementary Figure 30**). It is also observed that the frequency of these downward spikes increases with the length of peptide or linker. (**Supplementary Figure 31**). However, these spikes, which appear significantly different from tPAL events, would not interfere with the measurement."*

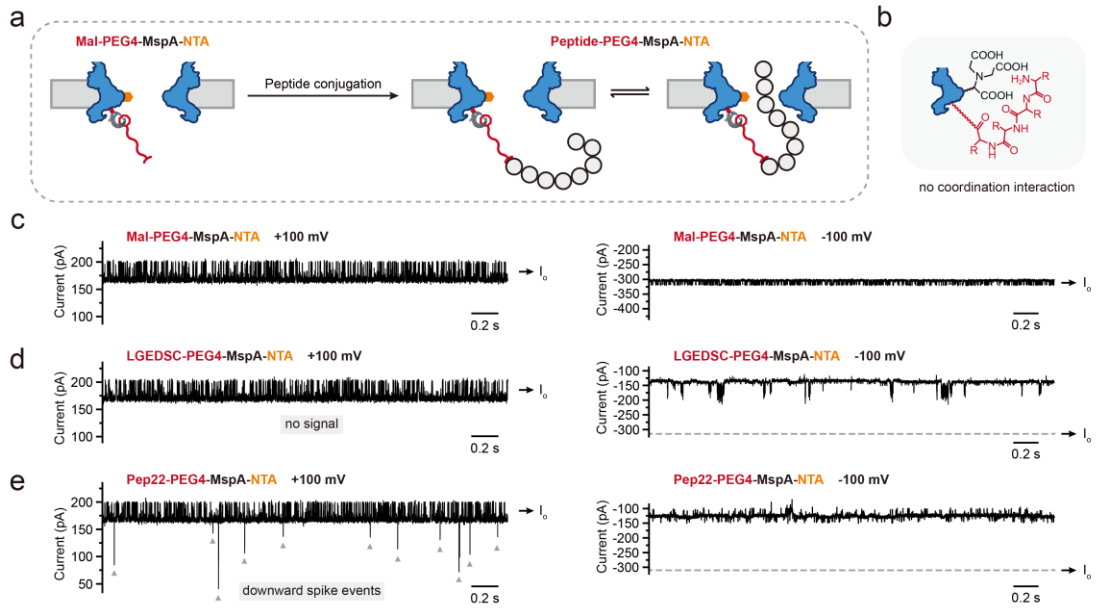


Figure RL 3.10. tPAL sensing of peptide without coordination interaction. (**a**) The construction of Peptide-PEG4-MspA-NTA. A cysteine-containing peptide is linked to Mal-PEG4-MspA-NTA via Michael addition, generating Peptide-PEG4-MspA-NTA. The immobilized peptide could enter the pore without coordination interaction. (**b**) Molecular structure of Peptide-PEG4-MspA-NTA. No coordination interaction exists between NTA and the peptide. (**c**) Representative traces of Mal-PEG4-MspA-NTA acquired at +100 mV and -100 mV. (**d**) Representative traces of LGEDSC-PEG4-MspA-NTA acquired at +100 mV and -100 mV. No signal was observed at +100 mV, whereas a prolonged blockage event induced by peptide trapping was observed at -100 mV. (**e**) Representative traces of Pep22-PEG4-MspA-NTA acquired at +100 mV and -100 mV. Downward spike events were observed at +100 mV, suggesting transient translocation of the peptide. Prolonged blockage event was observed at -100 mV, demonstrating translocation of the peptide through the pore constriction.

Review Comments:

If PEG-peptide conjugates can reach the sensing region without coordination, why do kon and Kd values differ across amino acids? For example, cysteine has an unusually high kon.

Response:

We deeply appreciate Reviewer 3 for prompting this inspiring discussion regarding peptide binding kinetics. In response to the earlier inquiry, the PEG-peptide conjugate might reach the sensing region without coordination. This however only produce extremely short residing spiky events, which are generally ignored during analysis. All long residing events

with well defined event features were generated by coordination, thus reporting a dependence of the amino acid composite of the peptide N-terminus.

Regarding cysteine, there are fundamental differences between amino acid and peptide in terms of their coordination mechanisms (**Figure RL 3.11**). Specifically, an anchored cysteine coordinates with the NTA-Ni adapter through its α-amino and α-carboxyl groups, while an anchored peptide coordinates through the N-terminal amino group and adjacent amide bond. This difference in coordination mechanisms likely contributes to the higher k_{on} value observed for cysteine. Relevant discussions in the manuscript have been pasted below for the ease of your reference.

"the terminal amino-N and the amide-N of the immobilized peptide can form a five-membered chelate with the NTA-Ni adapter, creating a transient pore analyte loop…The repetitive binding and dissociation of the immobilized peptide with the NTA-Ni adapter during tPAL sensing facilitate multiple re-readings of the same peptide, allowing for precise recognition of its N-terminus."

*"Notably, cysteine exhibited the longest event dwell time and the shortest inter-event duration compared to other peptides, likely due to its distinct binding mechanism[46] (**Extended Data Figure 2j** and **Supplementary Table 5**)."*

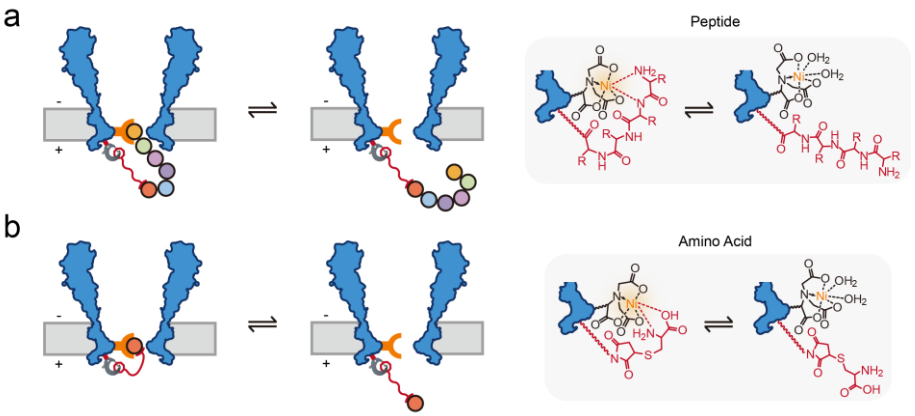


Figure RL 3.11. Different sensing mechanism of peptide versus amino acid with tPAL strategy. Amino acid primarily interacts with NTA-Ni adapter through α-amino and α-carboxyl group coordination, whereas peptide engages via N-terminal amino group and adjacent amide bond.

Review Comments:

The reported decrease in tPAL frequency with peptide or linker length raises the question of whether spike events become more frequent under these conditions.

Response:

We appreciate Reviewer 3 for raising this extremely insightful inquiry regarding the influence of peptide and linker length. The reviewer correctly noted that increasing either the peptide or linker length results in a reduced frequency of tPAL events and a concomitant increase in downward spike events (**Figure RL 3.12c, d**). Specifically, in the case of LGEDSC-PEG4-MspA-NTA-Ni, only characteristic tPAL sensing events were observed during the long-term recording at +100 mV. However, when peptide or linker was extended, downward spike events were detected in addition to tPAL events, as demonstrated by the sensing results for LGEDSC-PEG24-MspA-NTA-Ni and Pep22-PEG4-MspA-NTA-Ni (**Figure RL 3.12b**). As described in the previous section (**Figure RL 3.10**), these downward spikes represent transient blockages of the immobilized peptides. We attribute this phenomenon to the influence of spatial hindrance when the peptide or linker is longer. The spatial constraint makes it less favorable for peptide to access and form stable bonds with the adapter. As a result, the tethered peptide is more prone to transient interactions or blockages, which manifest as the observed spike events. This observation also highlights the importance of optimizing linker length in tPAL strategy.

We thank Reviewer 3 again for raising this insightful discussion. To fully address this issue, we have expanded the discussion (newly added **Supplementary Figure 31**) in the revised manuscript, pasted below for the ease of your references:

*"It is also observed that the frequency of these downward spikes increases with the length of peptide or linker. (**Supplementary Figure 31**). However, these spikes, which appear significantly different from tPAL events, would not interfere with the measurement."*

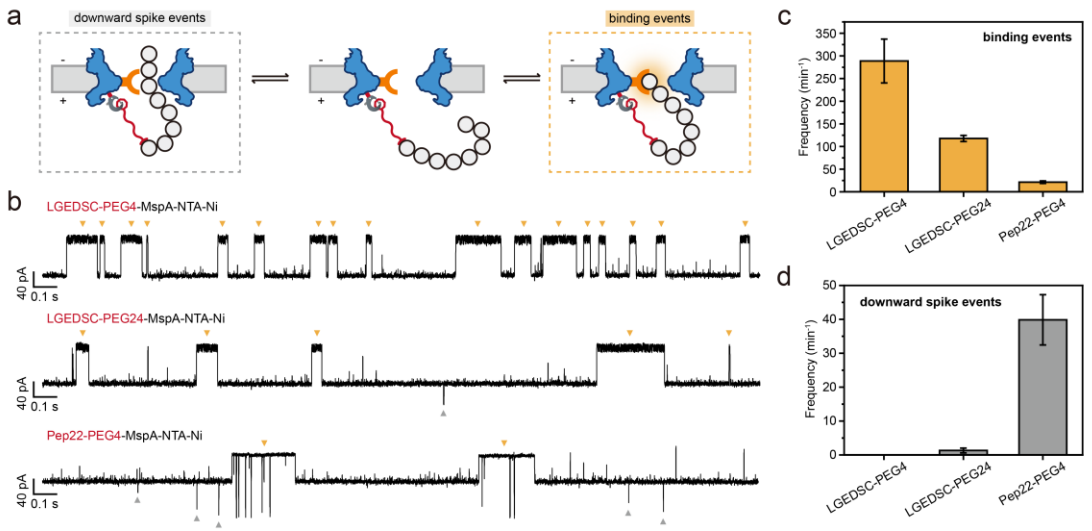


Figure RL 3.12. Comparation of events caused by immobilized peptide with or without coordination interaction. (**a**) Schematic diagram of tPAL sensing. When the immobilized peptide enters the pore and interacts with NTA-Ni adapter, a tPAL sensing event with well-defined features is observed. In contrast, when the peptide enters the pore without interacting with NTA-Ni adapter, only a downward spike event is detected. (**b**) Representative traces of LGEDSC-PEG4-MspA-NTA-Ni, LGEDSC-PEG24-MspA-NTA-Ni and Pep22-PEG4-MspA-NTA-Ni. All measurements were performed in a buffer of 1.5 M KCl, 10 mM CHES, pH 9.0 with a continually applied voltage of +100 mV. (**c**) Comparation of binding event frequency among LGEDSC-PEG4-MspA-NTA-Ni, LGEDSC-PEG24-MspA-NTA-Ni and Pep22-PEG4-MspA-NTA-Ni. (**d**) Comparation of downward spike frequency among LGEDSC-PEG4-MspA-NTA-Ni,

LGEDSC-PEG24-MspA-NTA-Ni and Pep22-PEG4-MspA-NTA-Ni. Generally, when the peptide length or the linker length increases, the frequency of binding events decreases but the frequency of downward spike events increases.

Review Comments:

Does peptide length affect the magnitude or fluctuations of the signal for a given residue? If so, this introduces a length-dependent complication; if not, the role of loop tension and conformation should be clarified.

Response:

We sincerely acknowledge Reviewer 3 for this insightful comment regarding the potential influence of peptide length on the event characteristics. As discussed in the previous section on spatial resolution, the event pattern of a given peptide is mainly determined by the combination of the first a few N-terminal amino acids and the impact of peptide length is relatively negligible.

To address this issue, we analyzed a periodic peptide sequence with repeated units (**Figure RL 3.13**). During stepwise enzymatic digestion of EFTREFTREFTRC, a periodic signal pattern was observed. Notably, within each repeating unit, the signals exhibited very similar characteristics when the same N-terminal residue was exposed, despite variations in peptide length. For example, aa1 (EFTREFTREFTRC), aa5 (EFTREFTRC) and aa9 (EFTRC) stages all produced events with the same low-noise features, However, subtle differences in current amplitude can also be observed between these similar events, attributed to variations in peptide length. Additionally, while the events at aa12 (RC) stage differed significantly from those at aa4 (REFTREFTRC) and aa8 (REFTRC) stages, this difference was primarily due to the large sequence differences between the peptides rather than their lengths.

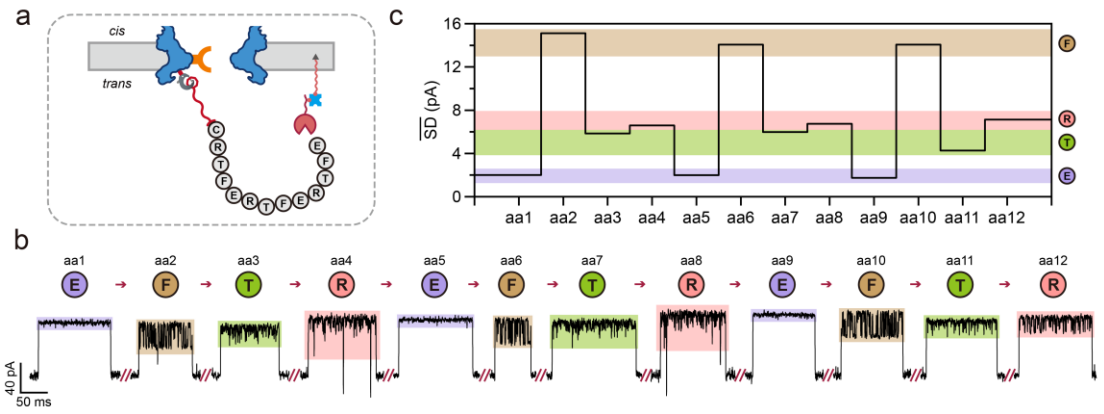


Figure RL 3.13. Sequential cleavage of EFTREFTREFTRC. (**a**) Schematic diagram depicting sequential cleavage of EFTREFTREFTRC using sgAP-dA15-chol. (**b**) Representative events obtained by sequential cleavage of EFTREFTREFTRC. The measurement was performed as described in **Methods**. Twelve enzymatic cleavages were completely observed during the real-time digestion process, producing thirteen stages with periodically repeating

pattern. To clearly illustrate this pattern, only the first twelve stages (aa1-aa12) are presented here. The corresponding N-terminal amino acid of each stage is marked above the trace. (**c**) The plot of $\overline{SD}$ at aa1-aa12 stages derived from EFTREFTREFTRC sequential cleavage.

Based on these experimental observations, we conclude that while the peptide length can have a minor influence on the signal magnitude, the primary determinant of signal characteristics is the composition of the N terminal peptide sequence.

We propose that, the peptide anchored via a flexible PEG linker experience relatively minor loop tension on its N-terminus. Therefore, the binding conformation is primarily dictated by the interactions between the peptide and NTA-Ni adapter, as well as the inner wall of the nanopore. This explains why peptides with similar N-terminal sequences exhibit similar signal features regardless of their length. This is further supported by the observation that peptides linked via PEG4 or PEG24 liner display highly similar signal patterns despite differences in linker length (**Supplementary Figure 15**). We anticipate that future integration of molecular dynamics simulations could provide a deeper understanding of loop tension and peptide conformation.

In summary, we acknowledge the importance of peptide length in influencing signal characteristics but emphasize that the sequence composition is the primary determinant of signal features. We have revised the manuscript to clarify this point and include additional data to support our conclusions. For the ease of reference, the added discussion is also pasted here:

*"Furthermore, for EFTREFTREFTRC, a peptide composed of repeating EFTR motifs, a distinct periodic signal pattern was detected (**Figure 5b, d**). Within each repeating unit, when the same N-terminal residue was exposed, the resulting events displayed similar noise characteristics with only minor differences in current amplitude, despite variations in their sequence length. The only exception was the aa12 stage, where the sensing events differed significantly from those at aa4 and aa8 stages due to differences in the downstream amino acid composition and the linkage to the pore."*

Review Comments:

Claims that tPAL events resemble stochastic sensing (Fig. S13) should be tested more broadly, and the contribution of loop tension explicitly considered.

Response:

We appreciate Reviewer 3 for this constructive suggestion. In the revised manuscript, we have expanded the comparison of sensing events between tPAL and stochastic sensing to explicitly clarify the contribution of loop tension (**Figure RL 3.7**). These results demonstrate that the similarity between tPAL and stochastic sensing events depends on peptide length. For longer peptides, the C-terminal linker exerts a smaller influence on tPAL signal, thus

the sensing events in both tethered and untethered states are more similar. However, for short peptides, the contribution of C-terminal PEG linker to tPAL signal becomes more pronounced, leading to larger discrepancies between tPAL and stochastic sensing events.

Review Comments:

Signals from PEG24- vs. PEG4-linked peptides are similar (Fig. S14), raising concerns about whether linker length meaningfully affects conformation.

Response:

We sincerely acknowledge Reviewer 3 for the insightful comment regarding the influence of linker length on peptide conformation. Based on the similar events observed for observed for LGEDSC under PEG4 and PEG24 linkage conditions (**Supplementary Figure 15**), we believe that it is experimentally evident that the linker length would not significantly alter the conformation of the immobilized peptide during its interaction with the NTA-Ni adapter. The PEG linker is a highly flexible polymer chain, and in the tethered state, it is unlikely to impose strong constraints on the N-terminus of the peptide, as discussed previously in the analysis of the similarity between tethered and untethered peptide signals (**Figure RL 3.7**).

We sincerely thank Reviewer 3 again for raising this valuable discussion. To address your concern, we have added relevant discussions of this point in the revised manuscript.

"The elongation of PEG linker didn't significantly alter its binding conformation and the produced event features."

Review Comments:

In addition, Fig. S15a shows unexpectedly flat signals after binding, which requires explanation.

Response:

We sincerely appreciate Reviewer 3 for prompting this insightful discussion on results of Fig. S15. According to our previous findings, high-fluctuation noise of sensing events is likely attributed to conformational changes of the analyte-adapter complex (Nature Methods, 2024, 21, 92-101). However, for 56Peptide-dT15-MspA-NTA-Ni, the pore constriction simultaneously accommodates the NTA-Ni adapter, an oligonucleotide and a peptide when in the bound state (**Figure RL 3.14a**). We speculate that this spatial constraint limits the motion of the peptide-adapter complex, forcing it to adopt a more restricted conformation during their interaction, which consequently results in flat signals after binding.

To fully address your inquiry, we further tested two additional peptides, LGEDSC and RGEDSC, under the same sensing mode. As shown in **Figure RL 3.14b-d**, both LGEDSC and RGEDSC produced low-noise binding events, similar to those of DGEDSC. This result further confirms our hypothesis.

This demonstration however indicates the low resolution of this sensing configuration with the peptide tethered to site 56 of the pore (pore rim). Therefore, throughout this manuscript, we adopted site 93 (to the *trans* side of the pore constriction) for peptide conjugation, which offers higher resolution. Relevant discussions have been added to the revised manuscript and are provided below for your reference.

"However, peptide immobilization to site 56 significantly constrains the conformation of the peptide-adapter complex, resulting in flat binding events."

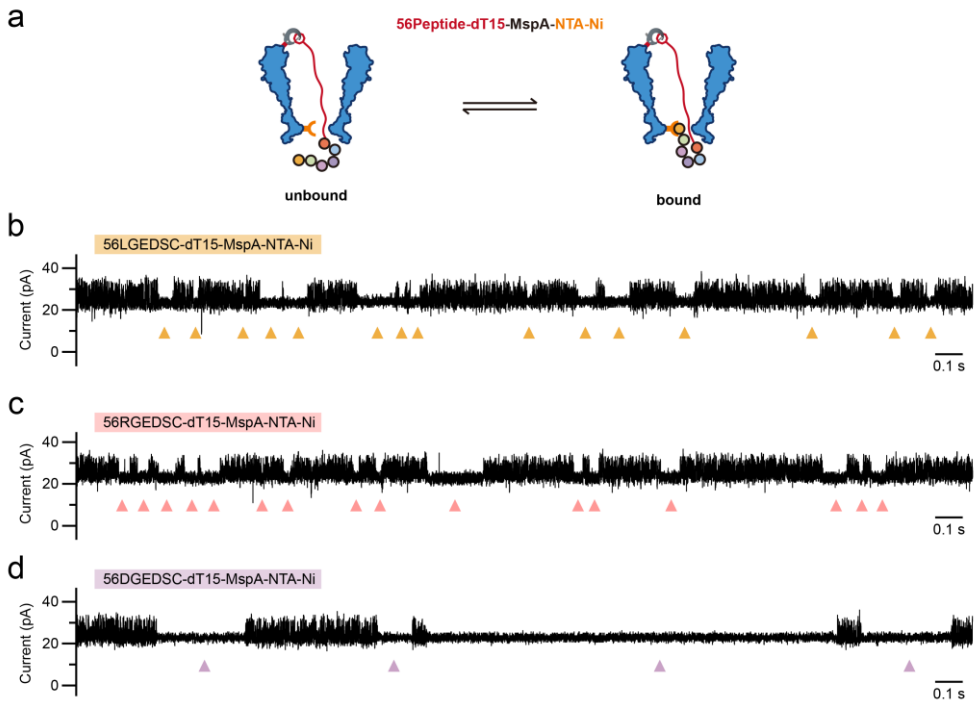


Figure RL 3.14. tPAL sensing of peptides conjugated at site 56. (a) Sensing mechanism of 56Peptide-dT15-MspA-NTA-Ni. (**b-d**) Representative traces respectively acquired with (**b**) 56LGEDSC-dT15-MspA-NTA-Ni, (**c**) 56RGEDSC-dT15-MspA-NTA-Ni and (**d**) 56DGEDSC-dT15-MspA-NTA-Ni. All measurements were performed in a buffer of 1.5 M KCl, 10 mM CHES, pH 9.0 and with a continually applied voltage of +140 mV.

Review Comments:

Differences in koff values (Fig. S21) suggest more complex peptide configurations than described.

Response:

We highly acknowledge Reviewer 3 for raising this valuable discussion regarding the dissociation rate (k_{off}) across different peptide. For tPAL sensing with NTA-Ni adapter, the dissociation kinetics are predominantly governed by the coordination interaction between the immobilized peptide and NTA-Ni adapter. This process is indeed influenced by peptide sequence and the binding configuration.

With respect to peptides in the original **Supplementary Figure 21** (renumbered as **Supplementary Figure 32** in revision), the difference in k_{off} between the two is not significant. Specifically, the k_{off} value for Pep14 is 4.89 ± 0.27 s^{-1}, while that for Pep16 is 3.84 ± 0.44 s^{-1} (**Figure RL 3.15d**). Notably, the sequences of these two peptides are identical from the N-terminus to cysteine residue, and their binding events are also very similar (**Figure RL 3.15a-c**), indicating that both peptides adopt similar conformation when bound to NTA-Ni adapter. However, Pep16 possesses two additional glycine residues following cysteine residue, which may account for the subtle difference observed in k_{off} values. To fully address your inquiry, we have added relevant discussions in the revised manuscript. For the reviewer's convenience, they are also provided below:

*"The binding events observed for Pep16 were comparable to those of Pep14 (**Supplementary Figure 32**), as both peptides share identical N-terminal sequences and only subtle differences in dissociation kinetics were observed."*

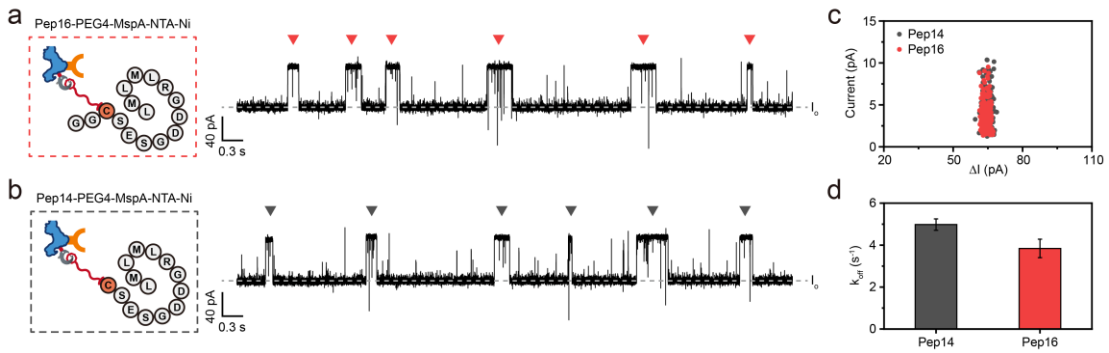


Figure RL 3.15. Comparation of tPAL events acquired with Pep16 or Pep14. (**a, b**) Schematic diagrams (left) and representative traces (right) for (**a**) Pep16-PEG4-MspA-NTA-Ni and (**b**) Pep14-PEG4-MspA-NTA-Ni. Pep16 incorporates a cysteine residue within its peptide backbone, whereas Pep14 features a cysteine at its C-terminus. Despite their differing cysteine positions, due to their identical N-terminal sequences, Pep16 and Pep14 exhibited almost identical tPAL sensing events. All measurements were conducted in a 1.5 M KCl, 10 mM CHES, pH 9.0 buffer at a constant voltage of +100 mV. (**c**) Event scatter plot of ΔI versus SD for events recorded with Pep16-PEG4-MspA-NTA-Ni and Pep14-PEG4-MspA-NTA-Ni. (**d**) Comparation of k_{off} between Pep14 and Pep16. All data were derived from results of three independent measurements.

Review Comments:

Further experiments with residues such as histidine at positions beyond the N-terminus (e.g., position 2 or 3; Fig. S30) would clarify how many residues are simultaneously detected.

Response:

We appreciate Reviewer 3's valuable comments regarding the results of histidine-containing peptides. Strictly following your request, we tested a peptide with the sequence LHFTRSC, in which the histidine is located at the second position to the peptide N terminus. During tPAL sensing, this peptide generated two distinct signals corresponding to N-terminal coordination (type 1) and histidine side chain coordination (type 2). Spontaneous switching between type 1 and type 2 states could also be observed occasionally (**Figure RL 3.16a**).

During sequential cleavage of LHFTRSC, two distinct sensing events were observed after the cleavage of the first residue, similar to the observations in HLFTRSC sensing (**Supplementary Figure 41**). When histidine residue is further cleaved, only N-terminal coordination events are detectable (**Figure RL 3.16b**). Therefore, tPAL strategy can accurately identify the position of histidine residue.

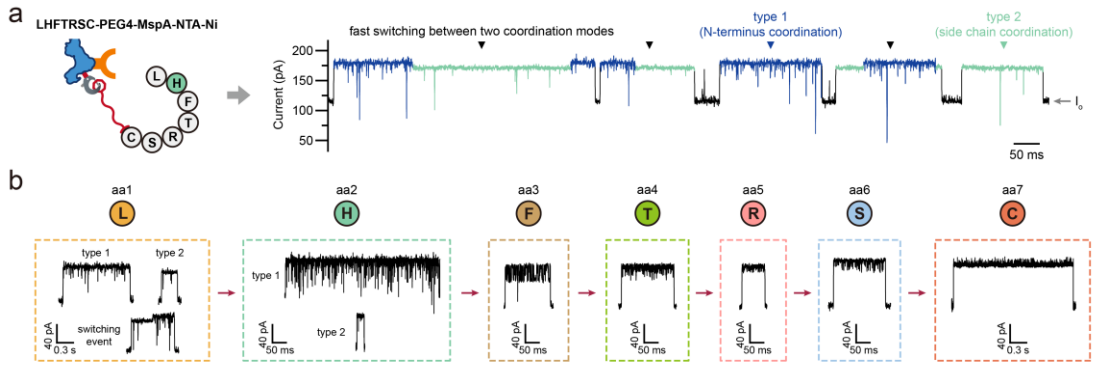


Figure RL 3.16. tPAL sensing and sequential cleavage of LHFTRSC. (**a**) Schematic diagram and a representative trace of LHFTRSC-PEG4-MspA-NTA-Ni. The measurement was conducted in a 1.5 M KCl, 10 mM CHES buffer at pH 9.0 with a continuously applied voltage of +100 mV. In addition to N-terminus coordination and side chain coordination events, another type of events generated by fast switching between the two coordination modes were also observed. (**b**) Representative events acquired during the sequential cleavage of LHFTRSC. The measurement was conducted using sgAP-dA15-chol as described in the **Methods**. Six enzymatic cleavages were observed during the sequential cleavage, yielding seven distinct stages.

With regard to the issue of spatial resolution, a detailed discussion has been already conducted in the preceding section (**Figure RL 3.8**).

Eventually, we again thank Reviewer 3 for your comment on the detection of histidine-containing peptide. Some relevant discussions and figure have been added to the revised manuscript to fully address this issue, as pasted below:

"Similarly, for peptides like LHFTRSC, where the histidine residue is located at the second position, both N-terminal and side-chain coordination events were observed, along with hybrid event resulting from rapid switching between the two coordination modes

*(**Supplementary Figure 42**)."*

Review Comments:

Minor Points

To illustrate changes in ionic conduction before and after chemical modification, all I-V curves should be plotted together for direct comparison.

Response:

Thanks Reviewer 3 for this insightful suggestion. Following your request, we have plotted all I-V curves together in the same figure (**Figure RL 3.17**). In this plot, the changes in ionic conduction caused by NTA modification and Ni^{2+} coordination are clearly distinguishable. However, the I-V curves of (N90NTA-Ni, N93AzK)$_1$(M2)$_7$ and Mal-PEG4-MspA-NTA-Ni are nearly identical, as the modification site of PEG linker is located outside the pore recognition, which does not significantly affect the pore conductance. To fully address your inquiry, a new Supplementary Figure has been added to the revised manuscript (**Supplementary Figure 4g**).

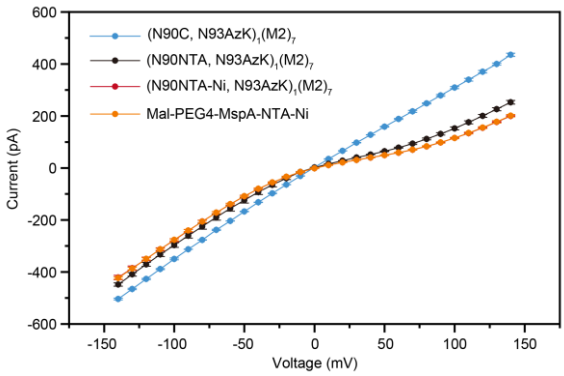


Figure RL 3.17. I-V curves acquired with (N90C, N93AzK)$_1$(M2)$_7$, (N90NTA, N93AzK)$_1$(M2)$_7$, (N90NTA-Ni, N93AzK)$_1$(M2)$_7$ and Mal-PEG4-MspA-NTA-Ni. A voltage ramp between -140 mV and +140 mV was applied to obtain the I-V curves. All statistics were derived from results of three independent measurements.

Review Comments:

The peptide modification process involves two steps: NTA-Ni^{2+} coordination at the N-terminus and cysteine-maleimide conjugation at the pore wall. Kinetic information on both steps, including the rate-limiting factor, should be provided to assess feasibility.

Response:

We appreciate Reviewer 3's thoughtful inquiry about peptide conjugation kinetics. As the reviewer correctly pointed out, the real-time conjugation process involves two sequential

steps. The mobile peptide introduced into *cis* chamber first enters the pore and interacts with NTA-Ni adapter via its N-terminus. This allows the C-terminal cysteine of the peptide to react with the maleimide group modified on the nanopore, thereby achieving anchoring.

To provide quantitative insights into this process, we employed LGEDSC for the demonstration. Upon addition of 20 μM LGEDSC, tPAL sensing events could be rapidly observed, with an average coupling time of 25 ± 8 seconds (n=3). We propose that the rate-limiting step in this process is the diffusion of the free peptides and their binding with the NTA-Ni adapter. Consequently, factors that influence the capture frequency and binding affinity, such as peptide concentration, composition, and net charge, will directly affect the conjugation kinetics.

To fully address your inquiry, the corresponding kinetic details have been added to the manuscript to clarify the feasibility of the peptide modification process.

"The average time required for the conjugation of LGEDSC at a final concentration of 20 μM is only 25 ± 8 seconds, and this can be further accelerated by increasing the peptide concentration. For other peptides, their capture frequency and binding affinity vary only slightly."

Review Comments:

While multiple rereads of a peptide are emphasized as a strength, they may also obscure resolution when residues are identical or generate similar signals. This limitation should be acknowledged.

Response:

We sincerely appreciate Reviewer 3's insightful comment regarding the potential limitation of the multiple rereads. We fully agree with your opinion that tPAL strategy struggles to achieve accurate identification for longer homopolymer sequences. However, from a proteomics perspective, long stretches of consecutive identical amino acids are extremely rare in natural proteins. For most peptides, the multiple reread strategy remains a significant advantage, as it can mitigate the interference of random noise and improve identification accuracy.

Additionally, as discussed in the section on spatial resolution (**Figure RL 3.8**), peptide sensing events are not solely determined by the N-terminal amino acid but are also influenced by subsequent residues. Therefore, for short repetitive sequences, tPAL method still holds potential for recognition. To illustrate this, we have conducted real-time enzymatic digestion measurement on AAAC, and successfully observed three cleavages. Although the first three amino acids of this peptide are all alanine, three distinct types of events were clearly and sequentially observed during the sequential cleavage process (**Figure RL 3.18**). Specifically, the peptide sequences in aa1 and aa2 stages are identical in first two amino

acids, and the resulting signals exhibit highly similar fluctuation patterns. However, due to the influence of the third residue, the event amplitude of aa1 and aa2 stages exhibits subtle differences sufficient for discrimination.

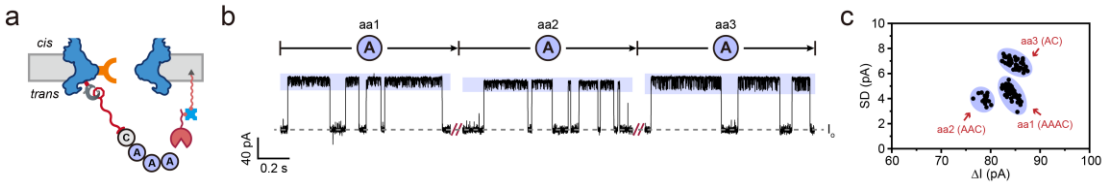


Figure RL 3.18. Sequential cleavage of AAAC. (**a**) Schematic diagram depicting AAAC sequential cleavage using sgAP-dA15-chol. (**b**) A representative trace acquired during sequential cleavage of AAAC. The measurement was performed as described in **Methods** using sgAP-dA15-chol. (**c**) The scatter plot of ΔI versus SD for events acquired at different stages described in (**b**).

We also acknowledge that different peptides may generate similar signals, making differentiation challenging. To address this issue, we have proposed a novel recording strategy based on multi-voltage sweeping in this revision. By leveraging the unique feature of tPAL, which ensures peptide immobilization within the nanopore throughout the measurement, this method enables repetitive sampling of the identical peptide under diverse voltage conditions. Experimental results demonstrate that while certain peptides may exhibit similar signals at specific voltages, differences can be observed in other voltages, thereby enabling clear discrimination. Utilizing this strategy, we successfully differentiated 20 XTRSC peptides with a machine learning accuracy of 98.0% (**Extended Data Figure 1**). For instance, although FTRSC and YTRSC show only minor signal differences at +100 mV and +140 mV, their signals are fully distinguishable at +60 mV (**Figure RL 3.19b**). By integrating sensing data across the three voltage conditions, the true positive rate (TPR) for FTRSC reached 100% (**Figure RL 3.19a**).

We sincerely appreciate the constructive feedback provided by the reviewer, which could help clarify the applicability of our methodology. In the revised manuscript, relevant discussions regarding the discrimination of repetitive sequences and similar signals are included. They are also pasted below for the ease of your reference:

*"Given that tPAL events are not solely determined by a single N-terminal amino acid, this method can also serve to analyze short homogenous sequences (**Supplementary Figure 57**), as demonstrated with the peptide AAAC. It however would have difficulties to decode long homogeneous sequences."*

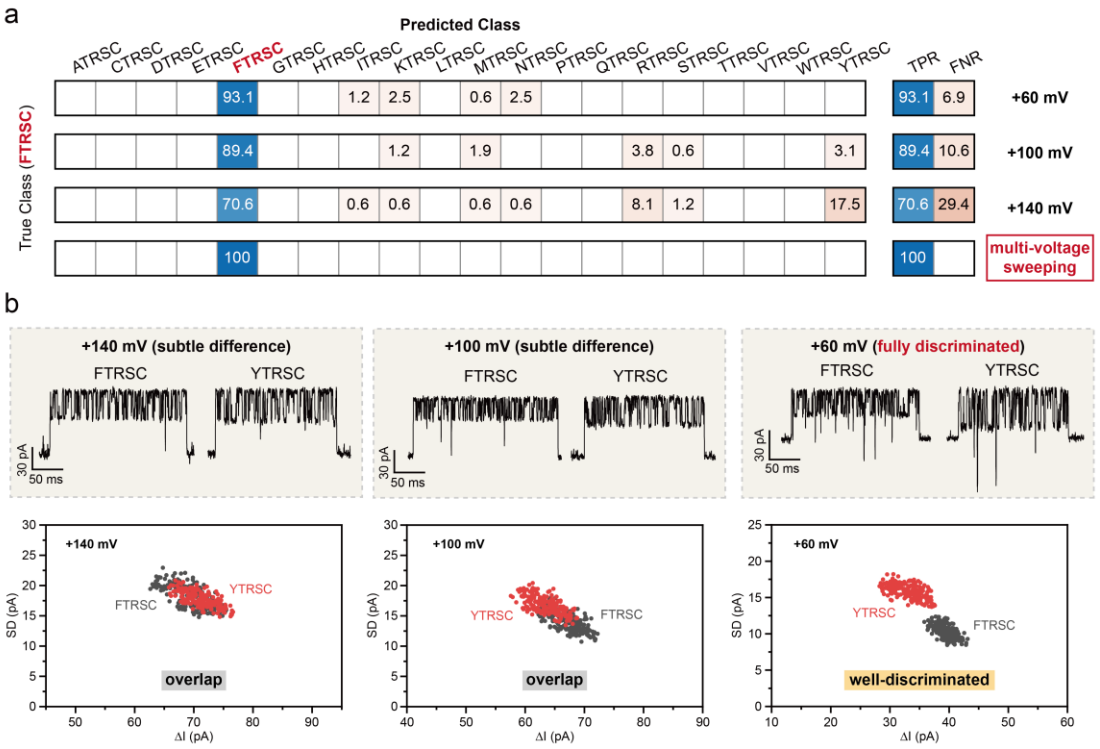


Figure RL 3.19. Machine learning performance improvement with multi-voltage sweeping strategy. (**a**) The improvement of the true positive rates (TPR) for FTRSC prediction with multi-voltage recording. (**b**) Top: Representative tPAL events acquired with FTRSC and YTRSC at +140 mV, +100 mV and +60 mV. Bottom: The scatter plots of ΔI versus SD for tPAL events acquired with FTRSC and YTRSC at +140 mV, +100 mV and +60 mV. While the event distributions have some overlap at +140 mV and +100 mV, they are well-discriminated at +60 mV.

Review Comments:

The apparent enhancement of sgAP activity arises from immobilization on the lipid membrane, which increases local concentration, rather than intrinsic activity changes. This should be clarified.

Response:

We acknowledge Reviewer 3 for this valuable suggestion to explain the mechanism of aminopeptidase immobilization strategy. The reviewer is correct that the enhanced sgAP performance is a consequence of the increased local concentration acknowledging the sgAP enrichment on the lipid membrane. In the revised manuscript, strictly following your suggestion, we have added relevant discussions. For the ease of your reference, it is also pasted below:

*"This complex, referred to as sgAP-dA15-chol (**Figure 3a**), does not affect the intrinsic activity of the aminopeptidase. However, it increases the local enzyme concentration, thereby enhancing the cleavage efficiency. Additionally, sgAP-dA15-chol does not interact with the NTA-Ni adapter, thus avoiding interfering event acquisition (**Supplementary***

***Figure 35**).”*

Review Comments:

In line 353, the correct reference is Fig. 5d and h, not Fig. 6d and h.

Response:

We highly appreciate Reviewer 3 for picking up this figure citation issue. The mentioned issue has been well corrected in this round of revision. We have also performed a thorough check of all figure citations throughout the revised manuscript to ensure no similar discrepancies remain.

Finally, we would like to express our sincere gratitude to reviewer 3 for your careful reading and constructive comments. Based on these suggestions, we have refined the graphical presentation and description, clarified the technical mechanisms and validated the feasibility of real-world peptide sequencing. These revisions have significantly enhanced the clarity and depth of our work. We truly appreciate your insightful guidance and look forward to your further feedback.

Referee #4:

Review Comments:

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Response:

We sincerely appreciate Reviewer 4 for your effort reviewing this manuscript. Your support to the publication of this work is deeply acknowledged.

Response Letter

Reviewer #1:

Review Comments:

The authors have responded comprehensively to my initial comments and have substantially strengthened the manuscript through the addition of new experimental data and improved technical clarification.

The additional SDS-PAGE methodological details are appreciated and improve reproducibility. However, clarification is needed regarding the electrophoresis conditions. The manuscript states that a 10% SDS-PAGE gel was run at 160 V for 16 h, which seems unusually long at that voltage and may represent a typographical error. Please clarify the actual running conditions used. In addition, please specify the SDS-PAGE buffer system employed (e.g. Tris-glycine-SDS or Tris-tricine).

Response:

We are delighted that our responses have comprehensively addressed all your previous comments. Regarding the SDS-PAGE methodology, we would like to clarify that the electrophoresis was indeed performed at 160 V for 16 h on the 10% gel. This extended running time was necessary to achieve complete separation of different pore assemblies with subtle differences. And the buffer system used here was Tris-Glycine-SDS. To fully address your concern, we have revised the **Methods** section to explicitly clarify the SDS-PAGE methodological details. The relevant content is also pasted below for the ease of your reference:

"The electrophoresis was run at 160 V for 16 h using Tris-Glycine-SDS running buffer."

Review Comments:

The effort placed into performing the suggested experiments to assess discriminative power of the system is appreciated. The revised manuscript now convincingly demonstrates discrimination of all 20 proteinogenic amino acids and utility of multi-voltage rereading to enhance resolution and classification accuracy. My concerns regarding dataset construction and statistical independence have been addressed by clarifying that machine-learning datasets were assembled from multiple independent pores. The revised terminology surrounding "de novo" sequencing is also more precise, and the scope of the decoding demonstration is now appropriately framed. Limitations related to throughput and scalability are acknowledged by the authors and are reasonably positioned as future engineering challenges.

Overall, the revisions largely address my prior concerns. Pending further clarification of points raised above, I would be supportive of publication.

Response:

We greatly appreciate Reviewer 1 for recognizing our efforts in the revision of this manuscript. All your constructive and inspiring suggestions have greatly improved the manuscript. We are truly grateful for your recommendation to publish this work.

Reviewer #2:

Review Comments:

Chen and coworkers have addressed all the questions I have raised. It is commendable to see how well the authors argued their points and how much extra work has been done to improve the manuscript. There is no doubt this work ranks highly among nanopore papers produced recently.

My concern, however, remains. as they are exclusively on the impact of this work on protein sequeincing. I think that journals of the calibre of Nature represent an important vehicle for science dissemination, not only reflecting the technical ability of the authors, but also the potential impact of the technology that is presented.

Therefore, I focus my comments exclusively on the innovative step, feasibility of this technology, on the challenges to make this technology a reality.

Response:

We are delighted that Reviewer 2 has recognized both the quality and extent of the additional work included in our last-round revision, as well as how well these efforts have addressed all of your previous comments. Your highly insightful and constructive feedback has undoubtedly helped further elevate the overall quality of this manuscript, and we sincerely appreciate your generous input. We are also deeply honored by your generous recognition of our work as one of the leading scientific studies in the nanopore field.

We also sincerely appreciate your insightful further comments in this round of review. It is worth noting that several of these points have already been addressed in our earlier revision (homopolymer measurement (original **Supplementary Figure 57**, now **Supplementary Figure 53**) and spontaneous insertion of peptide attached nanopores (original **Supplementary Figure 11**, now **Supplementary Figure 11** and **12**). We sincerely apologize for not clearly highlighting these results in the previous version of the manuscript, which may have led you to overlook these important points. Accordingly, in this round of revision, strictly following your guidance, we have supplemented more results and added further discussions to ensure that readers will not miss these key results. We have also expanded our validation of tPAL in the measurement of long peptide to address another suggestion raised by you. We believe that the revised manuscript has been further improved, and we again sincerely appreciate your time and effort in reviewing this work.

Regarding the broader implications of this work, while our demonstration primarily focuses on peptide sequencing, the tPAL strategy as a whole represents an entirely novel approach in the field. This innovation addresses a critical limitation: resolution constraints in nanopore-based peptide sequencing. It thus offers a promising but different pathway for more precise peptide sequence readout, significantly differing from the more prevalent

“strand sequencing” strategy.

Notably, we have also validated its versatility in handling complex scenarios, including diverse post-translational modifications (PTMs), chiral peptides, and peptides composed of unnatural amino acids. Additionally, preliminary investigations in our lab also suggest that the tPAL strategy may demonstrate efficacy in sequencing nucleic acids and glycans, further expanding its potential applications across molecular analysis in prospects.

Regarding your specific comments, our detailed responses are listed below in a point-by-point style, for the ease of the reference of you and the editorial team.

Review Comments:

Innovation.

It has long been demonstrated that nanopores can distinguish between molecules that have a similar size as the nanopore constriction. The authors have previously shown that MspA (the nanopore used here) can distinguish between all 20 amino acids and PTMs. Hence, it is not surprising that this approach can also do the same.

The novelty lies in attaching a peptide to the nanopore. This could allow single-molecule recognition of the terminal amino acid (see comment under ‘recognition’), but it brings many other issues, as underlinded below.

Response:

We sincerely appreciate your insightful observations regarding the innovative contributions of our work. While our prior research with the MspA-NTA-Ni nanopore platform achieved a landmark milestone in the field, first demonstrating the ability to distinguish all 20 canonical amino acids and their post-translational modifications, we would like to clarify several key distinctions that define this current work innovative:

- **Distinct Nanopore Architecture**: This study employs a substantially more advanced nanopore design, incorporating novel pore preparation techniques that represent a non-trivial improvement over the MspA-NTA-Ni. The new design requires significantly more complex preparation protocols, incorporating UAA by genetic code expansion (GCE) in a hetero-octameric MspA assembly.

- **The invention of tPAL:** Unlike previous work focusing on mobile amino acids, this research pioneers the analysis of tethered peptides immobilized within the nanopore environment. To enable the measurement, the N-terminus of the tethered peptide has to form reversible interaction with the reactive adapter designed at the pore constriction, a measurement configuration (tPAL) that has never been reported previously.

- **Fundamental Technical Distinction:** While amino acid identification is crucial, it is important to note that characterizing individual amino acids alone is insufficient to reconstruct complete peptide sequences. Although a proposed "chop and drop" approach in combination with nanopore amino acid detection might enable peptide sequencing in theory, the "chop and drop" strategy has not yet been experimentally validated or reported in the literature.

- **tPAL Mechanistic Innovation**: The tPAL method differs fundamentally from the "chop and drop" concept. Our approach could be conceptually defined as "sequencing by truncation" and has never been previously proposed in any prior review or perspective literatures, indicating the novelty of this idea.

- **Single Molecule Enzymatic Activity Monitoring of sgAP:** To the best of our knowledge, this should be the first single molecule observation of the enzymatic activity of sgAP. It is also fortunate to discover that the sgAP could work so well at a high salt buffer condition, making it a great companion for nanopore peptide sequencing.

- **Performance Advancements:** tPAL demonstrates superior sequencing capabilities compared to existing strand sequencing methods, with marked improvements in sequence consistency, sequence decoding, and single-amino-acid resolution.

- **Expanded Analytical Scope:** This platform also enables direct sequencing of peptides containing post-translational modifications (PTMs), chiral amino acids, and unnatural amino acid components, demonstrating a wide generality.

- **Future Applications:** Looking ahead, the tPAL methodology exhibits promising potential for adaptation to nucleic acid and glycan sequencing applications, opening new avenues for biomolecular analysis.

Finally, we appreciate your comments regarding the novelty of our work. We believe the innovation of this study is well supported by the concrete experimental results presented above, along with other evidences in the revised manuscript. The manuscript also provides comprehensive results that validate the consistency of our findings and include detailed analyses of event characteristics to further substantiate our claims. To improve clarity on our methodology, we have added two supplementary videos that explain the experimental procedures and show real-time measurements, an approach rarely adopted in the field, which we hope reflects our dedication to transparency. All machine learning code used in this work has been made publicly available in a user-friendly format.

As requested, in this round of revision, we have implemented the new experiments you suggested and incorporated the results into the revised manuscript. As a junior team,

publishing is never easy. Thus, we have invested significant effort in addressing all feedback with care to ensure rigorous scientific quality to earn us a precious opportunity of publication. Your support would be extremely valuable to us.

Review Comments:

Throughput.

This seems to me the major challenge. Peptides need to be attached to the nanopore and read once. Reversible chemistry with pre-inserted nanopores is required to obtain high thoughput, but it needs to be developed (and not using a disulfide bond strategy). It will alsol require buffer exchange to operate which will complicated the operations and increas operational time, and must be fast. If reversible chemistry cannot be developed, nanopores need to be reacted with the target peptides and re-reconstituted into membranes after every run. This would require many millions of nanopores simultaneously that record data with high resolution. Furthermore, it has to be demosrated that the nanopore with an attached peptide can insert (especially if the peptide is peptide, will allow insertion).

Response:

We sincerely appreciate Reviewer 2's insightful comments concerning the throughput. We fully recognize the critical importance of throughput for the practical application of the tPAL method in peptide sequencing. [redacted]

Notably, the MinION sequencing platform developed by Oxford Nanopore Technologies (ONT) already implements this modular architecture, demonstrating that the hardware feasibility has already been validated through existing commercial systems. Moreover, the development of nanopore acquisition systems capable of supporting thousands of parallel readouts has emerged as an industry-standard approach, with multiple commercial entities (including Oxford Nanopore Technologies, MGI, Qitan Technology, etc.) already operating such platforms at scale. These industry precedents confirm the technical viability of this architectural paradigm.

[Figure redacted]

We agree with you that the demonstration of spontaneous insertion of a peptide attached nanopore is critical. It is worth noting that this demonstration has already been provided in the original manuscript (original **Supplementary Figure 11**) but was unfortunately overlooked. We deeply apologize for not sufficiently emphasizing this in our original manuscript. To address this, in the revised manuscript, we now include results of a continuously recorded trace showing the moment of this spontaneous pore insertion. According to this result, upon pore insertion, the appearance of corresponding tPAL events can be immediately seen (**Figure RL 2.2**, new **Supplementary Figure 12**). Besides, we also further included results acquired with MspA conjugated with peptide of different sequence and length (SC or LMLMLRGDDGSESC) to show its generality (**Figure RL 2.3**, new **Supplementary Figure 11**).

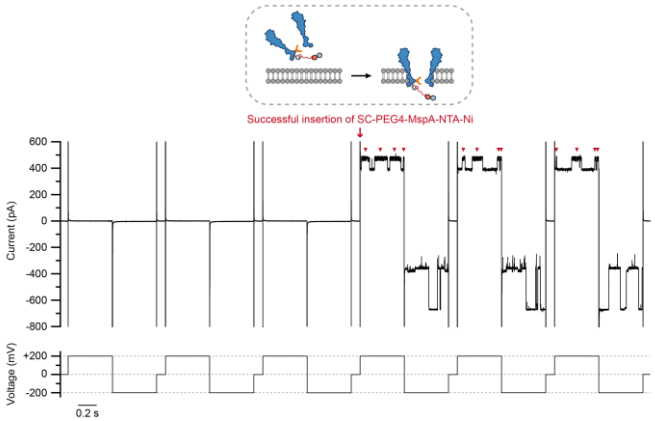
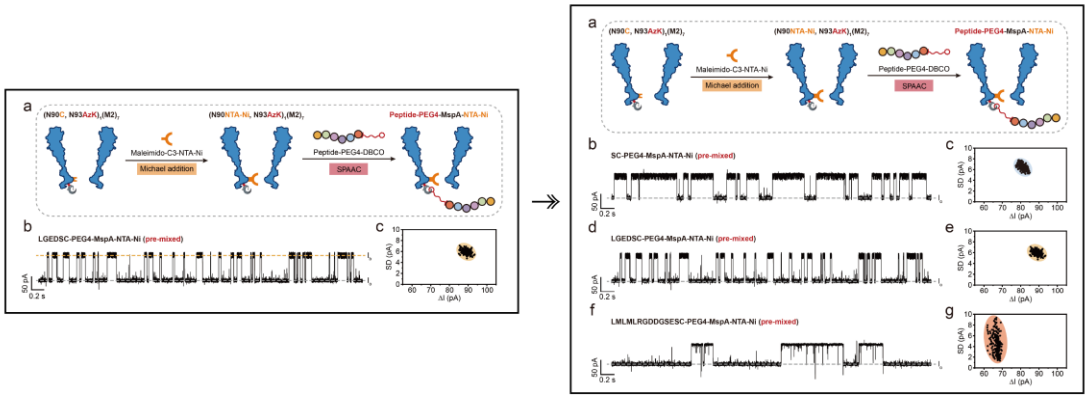


Figure RL 2.2. The spontaneous insertion of SC-PEG4-MspA-NTA-Ni. A programmed cycling voltage protocol (+200 mV for 0.5 seconds, -200 mV for 0.5 seconds, 0 mV for 0.1 seconds) was employed to trigger the insertion of peptide-conjugated nanopore.



original
Supplementary Figure 11

revised
Supplementary Figure 11

Figure RL 2.3. The original Supplementary Figure 11 and its corresponding modified version. Additional experimental results were incorporated to demonstrate the insertion capability of peptide-conjugated nanopore.

Eventually, we again thank Reviewer 2 for this constructive comment. The inclusion of above-mentioned result should fully address this issue. The added discussion is also pasted here for the ease of your reference:

*"A variety of pre-mixed Peptide-PEG4-MspA-NTA-Ni can also spontaneously insert into the membrane (**Supplementary Figure 11-12**)."*

Review Comments:

C-terminal tagging.

This is yet unsolved. The authors have rightfully indicated a few examples where C-terminal tagging has been used. However, that method is by no means generic or specific for all peptides. Lys-C tagging can also be used, but it would probably result in 50% of the peptides to be unusable, unless a lysine-specific tagging is identified. Nonetheless, I agree with the authors that this challenge might be eventually solved by the community at large.

Response:

We sincerely appreciate Reviewer 2 for raising this insightful discussion regarding C-terminal tagging. Our current work employs only a cysteine-based conjugation strategy to demonstrate its feasibility. We agree with Reviewer 2 that the development of a generalizable conjugation approach is essential.

In recent years, remarkable progress has been achieved in site-specific C-terminal peptide modification. Notably, the photoredox-mediated decarboxylative functionalization developed by the MacMillan group allows efficient and selective modification of the C-terminal carboxyl group in peptides (Nat. Chem. 2018, 10, 2, 205-211) and displays excellent substrate generality (**Figure RL 2.4**). This strategy has been successfully

implemented in fluorosequencing, where C-terminal alkyne modification was used for peptide immobilization (ACS Chem. Biol. 2021, 16, 11, 2595-2603). Furthermore, a recent study by the Cees Dekker group also adopted this functionalization approach to achieve C-terminal DNA conjugation, further demonstrating its broad applicability across diverse peptide sequences (J. Am. Chem. Soc. 2026, 148, 3, 3433-3441). Looking forward, this photocatalytic modification strategy can also be integrated into tPAL to enable C-terminal immobilization of a wide range of peptides. Collectively, based on the literatures cited above, we maintain that the methods discussed are indeed general and widely applicable (**Figure RL 2.4**).

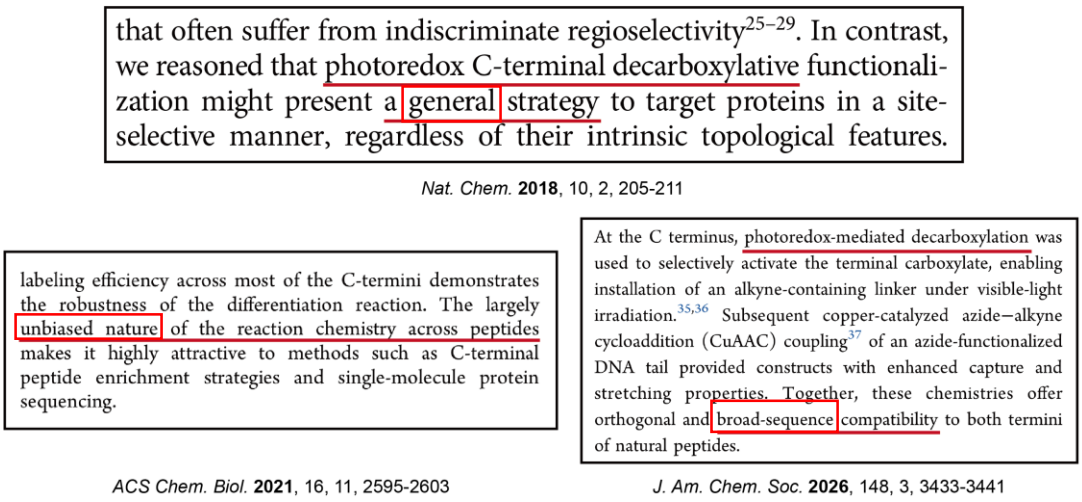
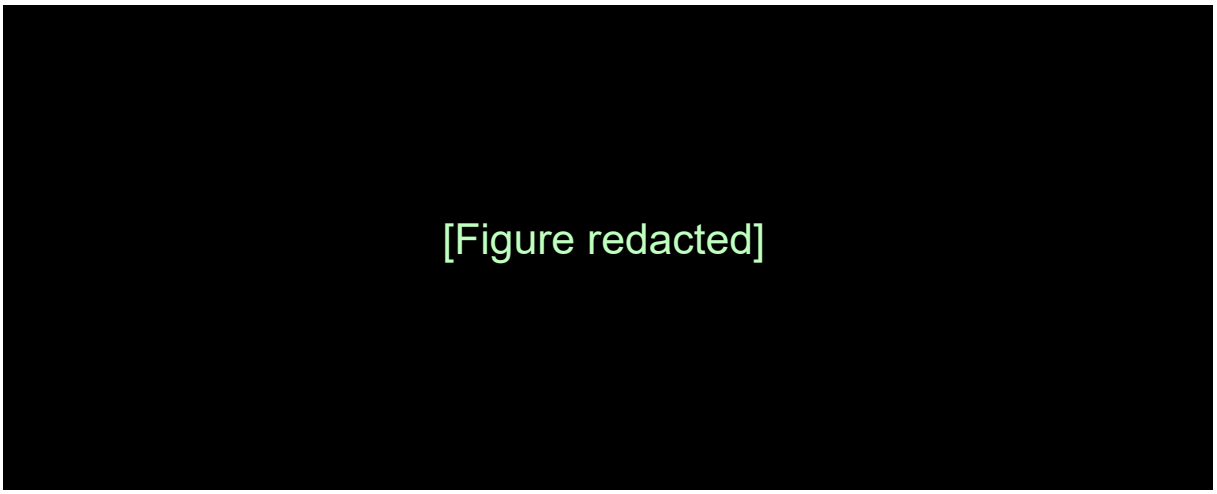


Figure RL 2.4. The universal applicability of the photoredox-mediated C-terminal functionalization method for peptides.

[redacted]

Notably, the peptide sequencing method reported by Quantum Si also employs a lysine side-chain anchoring strategy, which similarly encounters the issue of non-specific N-terminal conjugation. However, this issue becomes effectively negligible with high-throughput detection system. Additionally, considerable efforts are currently being directed toward the development of selective lysine modification tools to reduce the cross-reactivity with the N-terminus (J. Med. Chem. 2022, 65, 11840-11853; Angew. Chem. Int. Ed. 2022, 61, e202112107; Chem 2026, 12, 2, 102744), which can also be integrated into our tPAL method.

Eventually, we once again thank Reviewer 2 for your insightful comments regarding peptide-nanopore conjugation. As you noted, this challenge will ultimately be overcome through collective efforts across the community. Therefore, the tPAL strategy should be disclosed to the research community as soon as possible, so as to garner broader support and assistance for its further optimization and refinement. At this moment, your support would be extremely valuable to us.



[redacted]

Review Comments:

Recognition.

Several previous works have shown that 20 amino acids and their modifications can be identified by nanopore currents. The main authors have published recently the discrimination of 20+ amino acids using a very similar nanopore as indicated here.

The main issue is that the recognition is not of the last amino acid but of several amino acids at the constriction. The authors mention only three amino acids are recognised, but this is unlikely. Previous work with both DNA and proteins suggested several more amino acids are likely recognised, and the number of recognised amino acids will depend on their sequence.

But the main issue is the k-mer in combination with the short reads. Even if we assume that the peptidase will work in favour of sequencing (see next point), the combination of possible sequences is very large. For intact proteins passing through a nanopore, this issue is somehow mitigated by the length of the protein. Each amino acid is read multiple times (like for DNA sequencing) and the challenge is that of having enough information to identify one protein over a few thousand. Here, the number of possible combinations appears simply too large (multiples of millions). The authors have attempted building a simple classifier in the rebuttal, but it did not address the point.

Response:

We sincerely appreciate Reviewer 2's insightful commentary on the aspects of spatial resolution and sequence decoding. While it is accurate that our prior work established the nanopore identification capabilities of 20 amino acids and their PTM modifications via NTA-Ni functionalized MspA nanopores, this current study's core innovation extends far beyond the development of MspA-NTA-Ni, as articulated in the preceding section.

The tPAL approach fundamentally differs from conventional strand sequencing methods by employing precisely defined, reversible chemical interactions that impose strict conformational constraints on the peptide N-terminus. Unlike traditional methods that simultaneously detect signals from both N-terminal and C-terminal residues around the amino acids residing at the pore constriction, this approach report event features contributed from the peptide's N-terminus and few amino acids afterwards. The NTA-Ni adapter also provides chemical constraints which minimizes thermal fluctuations of the captured peptide so that a high measurement consistency was achieved. However, the conventional strand sequencing strategy also can't minimize thermal fluctuations of the target peptide strand within the pore constriction, generating even more fluctuation noises and uncertainties in sequence decoding. Moreover, the unique tPAL configuration enables repeated re-reading of the same peptide, further enhancing signal resolution. This capability collectively produces highly reproducible event profiles with well-defined characteristic signatures.

Through systematic investigations detailed in the revised manuscript (**Supplementary Figures 54-55**), we rigorously validated that amino acid substitutions beyond the third residue induce negligible signal perturbations. The data consistently demonstrate that event noise profiles are primarily governed by the first three amino acids, with downstream residues contributing diminishing signal features that rapidly attenuate with increasing distance from the N-terminus. Guided by these observations, our proof-of-concept sequence decoding successfully implemented a 3-mer model, achieving accurate sequence recognition without sacrificing practical applicability.

While larger k-mer models theoretically enhance recognition accuracy, a delicate balance must be considered between computational requirements, database size, and precision demands. Notably, the 3-mer decoding model requires just 8,000 sequence combinations instead of multiple millions. Experimentally, several amino acids, including glutamic acid (E), glutamine (Q), phenylalanine (F), threonine (T), and alanine (A), demonstrate distinct event features when positioned at the N-terminus of peptides, which could efficiently minimize the computation power required in future decoding algorithms. For instance, in most cases, N-terminal glutamic acid generates low-noise, distinctive signature profiles. Besides, when combined with rapid advances in AI algorithms, further improvements in data decoding algorithm performance can be further boosted. The ongoing development of ONT's nucleic acid sequencing capabilities serves as a representative example.

Review Comments:

Amino peptidase reading.

This approach requires that the 1) terminal peptide is cut one amino acid at a time, 2) the amino peptidase disengages from the peptide, 3) the terminal amino acid is read once, 4) the amino peptidase binds again to the terminus of the peptide and cuts it once more.

Previous work indicated that amino peptidases do not cut-disingage-cut-disingage… but they bind and then cut several peptides at the same time. However, let us assume an amino peptidase can be engineered to cut once and then disingage the peptide. In this case, there would still be a competition between nanopore readout and re-engaging with the peptidase. This could be minimised by having a very low peptidase activity, effectively reading the peptide many times. However, this would prevent reading two identical consecutive amino acids. I cannot see how to identify consecutive amino acids that are identical. The authors should try to design an experiment. Can peptides with stretches of, say, DD, DDD, and DDDD be identified? Maybe if the k-mer size is actually large, this difference can be identified.

Response:

We sincerely appreciate Reviewer 2 for the insightful comments regarding the "enzymatic cleavage mechanism" and "homopolymer recognition".

Regarding the enzymatic cleavage mechanism, we employed *Streptomyces griseus* aminopeptidase (sgAP), a well-characterized and widely applied aminopeptidase, for sequential cleavage of immobilized peptides. Our nanopore measurements unambiguously demonstrate that the amino peptidase operates in a distinctive "cut-disengage-cut-disengage" mode, appearing as sequential report of nanopore events corresponding to the sequence of the target peptide. Specifically, for a given target peptide, the observed stepwise signal changes precisely matching the peptide length during sequential cleavage of each peptide (for example, **Figure 3**). In different independent trials, identical stepwise signal changes were observed, further confirming its consistency (for example, **Supplementary Figure 38-39**). However, if the sgAP doesn't follow the "cut-disengage-cut-disengage" model, traces acquired from different independent trials won't exactly match each other. Besides, the newly added long peptide data (**Figure RL 7**, **Supplementary Figure 52**) serves as an additional compelling example. The provided a video clip of the real-time digestion process (**Supplementary Video 1**) should provide a vivid demonstration of this "cut-disengage-cut-disengage" activity.

While we are not ware of any prior studies specifically addressing the cut-disengage kinetics of sgAP, it is very likely that this newly developed nanopore platform provides the first real time single-molecule enzymatic activity of sgAP, further re-enforcing the innovation of this study. Experimentally, the observed kinetic profile consistently aligns with the "cut-disengage-cut-disengage" model. We further hypothesize that this enzymatic behavior may arise from a combination of experimental conditions, including: (1) immobilization of the peptide substrate in the vicinity of the nanopore constriction; (2) high-salt measurement conditions; (3) spatial enrichment of engineered sgAP near the lipid bilayer membrane; and (4) the presence of a strong electric field gradient at the pore constriction region. These factors collectively may have facilitated the "cut-disengage-cut-disengage" observed in our system.

We appreciate your raising of discussions regarding homopolymer analysis, which we acknowledge that it remains a persistent challenge in peptide sequencing. Indeed, as previously noted in nucleic acid sequencing literature (Nucleic Acids Research, 2025, 53, gkaf131), homopolymer identification presents inherent difficulties across sequencing technologies.

In fact, in our original manuscript, we already demonstrated tPAL's capability for short homopolymer characterization through sequential cleavage of AAAC (original **Supplementary Figure 57**, now **Supplementary Figure 53**). Our data revealed distinct signal patterns for AAAC, AAC, and AC, enabling unambiguous recognition of consecutive identical residues in the target peptide. We regret that this critical section was not emphasized adequately in the revised version, potentially leading to the oversight.

To address this concern even more comprehensively, we have re-analyzed a longer homopolymer motif AAAAC (**Figure RL 2.6**). While the event features of AAAAC and AAAC

exhibit close similarity, the cleavage-induced transition from AAAAC to AAAC produces a consistent amplitude decrease that remains visually and statistically distinguishable. This observation aligns with our fundamental mechanistic model: the first few residues establish the primary signal architecture, while subsequent amino acids produce gradually diminishing event feature contributions and a 3-mer data decoding algorithm already provides a satisfying sequence decoding performance. Therefore, we can still identify cleavage of an N terminal amino acid within a section of consecutive identical residues by leveraging the subtle but irreversible signal changes. Nevertheless, we acknowledge that this method would face significant challenges in resolving extremely long homopolymeric sequences. Fortunately, extended homopolymeric regions are relatively rare in natural proteins.

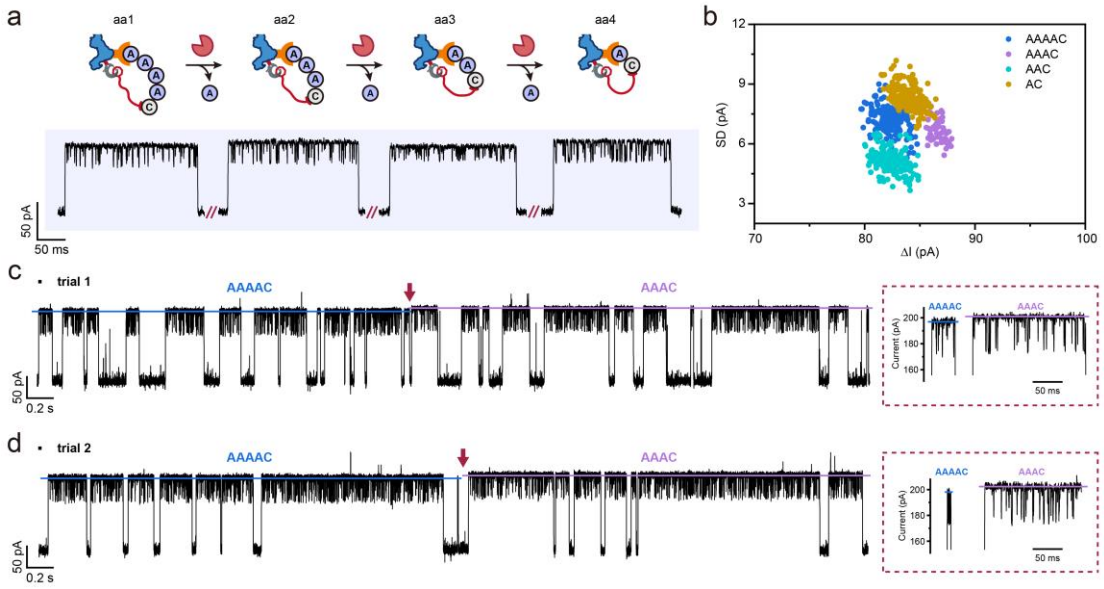


Figure RL 2.6. Recognition of homopolymeric sequence. (**a**) Representative events acquired during sequential cleavage of AAAAC. (**b**) The scatter plot of ΔI versus SD for tPAL events acquired from AAAAC sequential cleavage. (**c, d**) Left: The raw trace of tPAL event changes from AAAAC to AAAC. Right: The zoomed-in view of the area indicated by a red arrow in the left trace. A slight event amplitude increase was observed irreversibly during the transition from AAAAC to AAAC, marking the cleavage of a terminal A. This phenomenon is also validated in different independent trials, confirming its robustness.

Eventually, we once again thank the reviewer for your insightful and inspiring comments, which have promoted meaningful discussions on the enzymatic cleavage mechanism and homopolymer recognition. To address these, new experimental results and some relevant discussions are also added to the revised manuscript. These discussions are also pasted below for the ease of your reference:

*"Furthermore, given the high resolution of the pore, this method could also resolve short homopolymeric sequences, as demonstrated with the peptide AAAAC (**Supplementary Figure 53**). However, it may be challenging for tPAL to sequence peptide containing long homopolymeric segment."*

Review Comments:

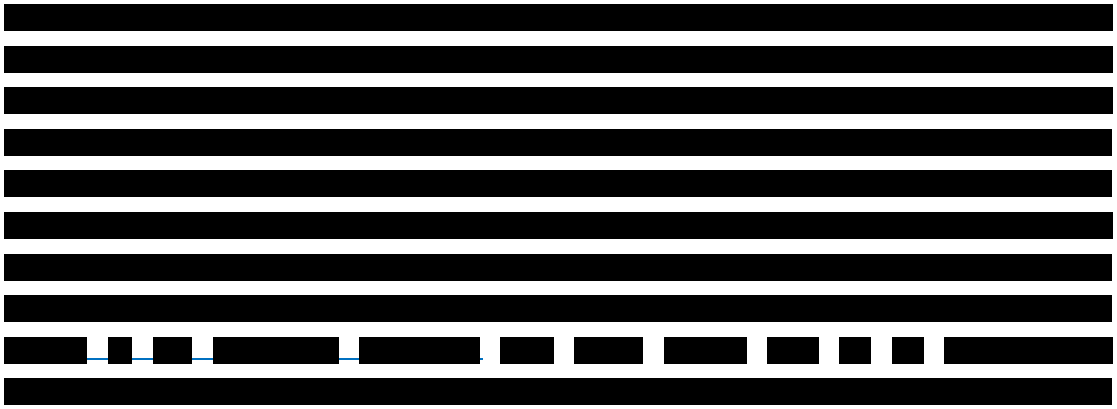
Peptide length.

The authors mentioned they used a peptide of 22 amino acids, but I could not find the data (Figure S28 shows the sequential cleavage of LRGESC, maybe they indicate S19? There there is a 22 amino acid peptide, but only three steps are observed. The authors mention that denaturing conditions can be used, but this would inactivate the enzyme. The structure of long peptides will affect the k-mer recognition. The authors should clearly state if they observe 22 cleaveage steps of Pep22.

Response:

We sincerely appreciate Reviewer 2 for raising this valuable point regarding long peptide detection. We would first like to clarify that the tPAL sensing results for Pep22 were indeed presented in **Supplementary Figure 28** of our previously-revised manuscript. We note that **Supplementary Figure 19** referenced by the reviewer corresponds to the numbering used in our initial submission. We believe this discrepancy may have arisen from the use of figure labels across different manuscript versions. We kindly suggest the reviewer to consult our latest revised manuscript for the most up-to-date and consistent figure numbering and corresponding results.

The tPAL method theoretically enables detection of any peptide containing an exposed N-terminus regardless of molecular length. This analytical capacity was empirically validated through successful detection of the 22-amino acid peptide (Pep22). However, comprehensive cleavage of Pep22 at aspartic acid (D) residues remains significantly hindered by the current aminopeptidase's sequence-specific kinetic limitations. Consequently, the corresponding cleavage data was not shown. These limitations are not inherent to the methodology itself but rather reflect the enzymatic constraints of the current reagent system. Future advancements through enzyme engineering strategies or implementation of synergistic enzyme cocktails are anticipated to resolve these sequence-dependent inefficiencies, thereby enhancing the method's universality for diverse peptide substrates through optimized hydrolytic performance.



We agree with Reviewer 2 that the tPAL methodology may be limited in detecting peptides that fail to have an exposed N-termini, for example, highly structured long peptides. We also concur with the reviewer's concern regarding potential enzyme activity compromise when using denaturants to achieve peptide structural unfolding. Fortunately, both the MspA nanopore and the sgAP aminopeptidase demonstrate exceptional structural stability, as evidenced by extensive prior investigations on MspA's conformational resilience (Chem. Sci., 2021,12, 9339-9346; J. Biol. Chem., 2003, 278, 8678-8685). Furthermore, the sgAP aminopeptidase exhibits remarkable thermal and chemical stability, retaining enzymatic activity at near 70°C and in high-salt buffers - conditions that typically inactivate many enzymes. These properties suggest that sgAP may possess inherent denaturation resistance, forming the basis of our current hypothesis.

The critical scientific challenge remains optimizing denaturant concentrations to maintain equilibrium between efficient peptide unfolding and preserving enzymatic functionality. Our future research agenda includes comprehensive stability evaluation of sgAP under varying denaturing conditions. We also anticipate that AI-driven enzyme evolution will enable structural modifications to enhance sgAP's stability against denaturant, potentially expanding the detection capabilities for structurally complex peptides. However, it is also likely that the sgAP is already stable against weak denaturant. Alternatively, we propose enzymatic hydrolysis of long peptides into shorter detectable fragments for tPAL analysis, thereby minimizing denaturant dependency.

We sincerely appreciate Reviewer 2 for this constructive insight that has strengthened our scientific discourse. To fully address your comment, the following comprehensive discussion has been incorporated into the revised manuscript and is pasted below for the ease of your reference:

*"Another peptide consisting of six EQTR motifs (EQTREQTREQTREQTREQTREQTRC) also report periodic signal patterns directly correlated with its peptide sequence (**Supplementary Figure 52**), confirming the feasibility of tPAL for long peptide (>25 aa) sequence analysis."*

"A read length of up to 25 amino acids was achieved, with longer reads also deemed feasible but not experimentally examined herein."

Review Comments:

In conclusion, despite the best effort from the authors to improve and explain their method, I maintain the view that it appears extremely unlikely this approach will allow peptide sequencing.

The main issues I believe the authors should address are:

1. It has not been demosntrated that long peptides can be addressed by this approach - i.e. the authors should show that the 22mer peptide can originate 22 amino acid steps (not necessarialy sequenced)

2. Homopolymeric sequence should not be possible if the k-mer is small (an experiment can be designed for studing this)

It should be noted that if the k-mer reads more than 3 amino acids, homopolymeric peptides should become less of an issue. However, peptide sequencing would become very challenging given the astronomical possible combinations of peptides. However, it would not be impossible, as in theory the sequential cleveage by the enzyme would allow to read individual amino acids multiple time.

Response:

We are deeply grateful to Reviewer 2 for your detailed and professional evaluation of our manuscript. Your thoughtful insights on key aspects including throughput, C-terminal tagging, enzymatic cleavage mechanisms, long peptide and homopolymer measurement, have all further improved this manuscript. Through multiple rounds of meticulous revision, we are confident that these concerns have been comprehensively resolved by concrete experimental evidences. The revised manuscript now presents a robust demonstration of the tPAL sequencing approach's feasibility, with ongoing developments showing strong potential to broaden its applicability in proteomics study.

Our sincere appreciation is also extended for your encouraging comments and continued support. This work introduces a novel peptide sequencing strategy that demonstrates distinct advantages over conventional methodologies, offering innovative solutions for long-standing technical barriers in the field. We hope this revised version effectively conveys both the scientific merit and practical promise of the tPAL method. If you have any additional questions or require further clarification, we would be honored to provide supplementary information at your request.

Reviewer #3:

Review Comments:

I have reviewed the revised version and the associated responses. Relative to the prior version, the manuscript has been substantially expanded in experimental scope and analysis. The authors present (i) a dual-functionalized hetero-octameric MspA pore enabling peptide immobilization and repeated N-terminal interrogation via NTA-Ni coordination (“tPAL”), (ii) stepwise enzymatic cleavage to generate sequential, stage-dependent signal changes, and (iii) broader demonstrations including PTMs (notably lysine acetylation vs trimethylation), phosphorylated substrates (with an enzymatic workaround), unnatural amino acids, D-peptides (sensing only), repeating-sequence peptides, and a machine-learning pipeline for stage classification and proof-of-concept sequence reconstruction. Overall, the revised work is data-rich and more convincing technically than the original submission. I recommend acceptance.

Response:

We highly appreciate Reviewer 3 for considering our revised manuscript data-rich and technically convincing. We truly appreciate the time and effort you have dedicated to reviewing our manuscript and for providing thoughtful comments that have significantly improved the work. Your recommendation to publish this work is highly appreciated.

Reviewer #4:

Review Comments:

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Response:

We sincerely appreciate Reviewer 4 for your effort reviewing this manuscript. Your support to the publication of this work is deeply acknowledged.

Response Letter

Reviewer #2:

Review Comments:

The authors have demonstrated excellent argumentative skills, thoroughly addressing each comment. As previously mentioned in my review, however, my evaluation goes beyond the technical assessment of this work or the argumentative skills of the authors, but rather focuses on the novelty of their approach and its applicability to protein sequencing.

As previously noted, numerous nananopore experiments have demonstrated the ability to distinguish between individual amino acids. Furthermore, unnatural amino acids have been introduced into nanopores, and the MspA-NTA-Ni nanopore has been utilized in numerous articles authored by the main author.

Therefore, the innovative step involves attaching a peptide to the nanopore and reading the initial amino acid through current recording.

Response:

We sincerely appreciate the reviewer for the continued engagement and for acknowledging that ALL previous concerns have been thoroughly addressed. In the same spirit and to the respect of the reviewer, we have approached each point in the previous round with the same commitment to providing direct and concrete experimental evidence, accompanied by rigorous and comprehensive discussion. Furthermore, to ensure clarity, in each round, we have deliberately included detailed comparative figures and clarifications within the response letters, for the ease of all reviewers and the editorial team members.

Regarding the novelty, we wish to emphasize that our central innovation lies not in the mere conjugation of peptides to nanopores, but in the conceptual development and experimental realization of the tPAL method and it is the first work in this series of study. Here, tPAL enables: 1) specific molecular moieties (demonstrated here but not restricted to peptides) to form reversible covalent/coordination bonds with chemically engineered reactive adapters (demonstrated here but not restricted to NTA-Ni), and 2) repetitive measurement of the same single molecule via dynamic binding/unbinding processes. This unique configuration enables continuous temporal tracking of any subtle chemical changes of the same analyte (demonstrated here but not limited to peptide shortening). We also demonstrated enzymatically removal of phosphorylation accompanied with exopeptidase, in an enzyme cocktail fashion (**Supplementary Figure 51**), a specialized application of tPAL, also demonstrating real time monitoring of subtle chemical changes of a tethered analyte and the versatility of the method itself.

This innovative approach necessitated the development of a dual-functionalized nanopore platform with enhanced detection resolution. We thus developed the first heterogeneous MspA nanopore engineered with unnatural amino acids, achieving completely orthogonal bifunctional pore modification unprecedented in the field. Notably, the preparation of heterogeneous MspA itself is technically demanding, and incorporating unnatural amino acids (UAAs) into such heterogenous pore constructs presents even greater challenges, which might explain why this method has yet to be reported in the literature. It thus differs fundamentally from established architectures like MspA-NTA-Ni. Notably, as confirmed through comprehensive patent and literature searches, we have developed and characterized a nanopore construct with no prior precedent in the literature.

Given the central importance of peptide sequencing in the field, we have prioritized this demonstration with peptide sequence analysis to establish the method's transformative potential in the context of conventional sequencing limitations. In principle, by different combinations of reactive adapters, target analytes and the enzymes used, tPAL is also capable for analysis of glycan, nucleic acid, and post-translational modification (PTM) detection. Preliminary data supporting these applications are also demonstrated to the editor. The preliminary demonstration with peptide phosphorylation using an enzyme cocktail approach (**Supplementary Figure 51**) has also been included in the manuscript as well.

This approach also exhibits exceptional spatial resolution, which overcomes the long-standing difficulty of resolution inherent in conventional strand sequencing methods. With its high performance, tPAL represents a significant technological advance in the field of nanopore-based peptide sequencing and clearly differs from the strand sequencing strategy. Timely sharing of this groundbreaking technology with the broader academic community would collectively advance progress in peptide sequencing, and we are sincerely grateful for your support in enabling this collaborative endeavor.

Review Comments:

Nevertheless, I believe this approach presents several practical challenges:

1. The peptide must be attached to the nanopore, and the nanopore-peptide must be inserted into a membrane. While the authors have demonstrated that small peptides allow insertion, it is uncertain whether large peptides will be tolerated. This has not been substantiated by additional experiments.

Response:

We sincerely thank Reviewer 2 for the comments regarding the insertion ability of peptide attached nanopore. Actually, ever since our original manuscript, we have already demonstrated insertion of nanopore with conjugated peptide. Specifically, in our previous

round of revision, we have already provided detailed experimental evidence demonstrating that the MspA nanopore conjugated with a 14-mer peptide could insert into the membrane for tPAL measurements (**Figure RL 2.1**), which clearly confirms the compatibility of large peptides with the tPAL system. We deeply regret that this critical point was overlooked by the reviewer despite its explicit inclusion in the supplementary figures and detailed textual explanation.

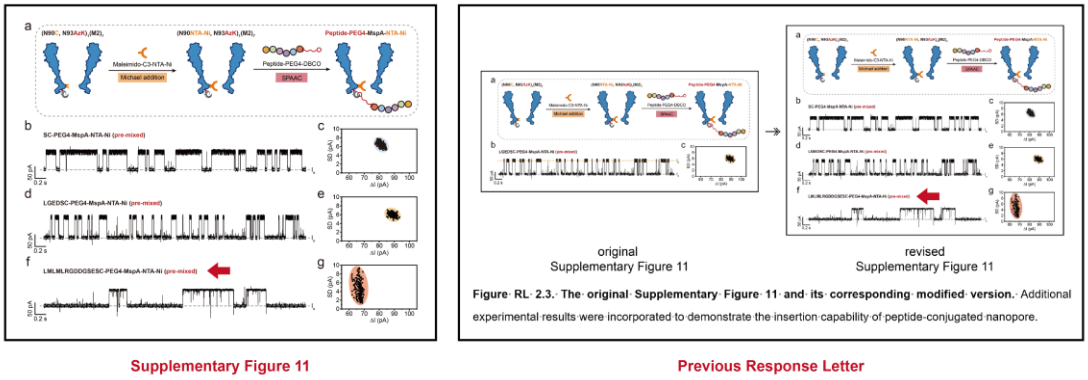


Figure RL 2.1. The demonstration of pore insertion ability with 14-mer peptide in our previous revision. Left, the revised manuscript. Right, the description in the previous response letter.

To strictly follow reviewer 2's additional request, we have now included concrete evidence demonstrating spontaneous insertion of the 22-mer peptide attached MspA nanopore (**Figure RL 2.2**). This result fully validates the spontaneous insertion of nanopore with a conjugated long peptide (22 mer). We have also added a schematic illustration above the representative trace to improve clarity and guide readers through the key features of the data.

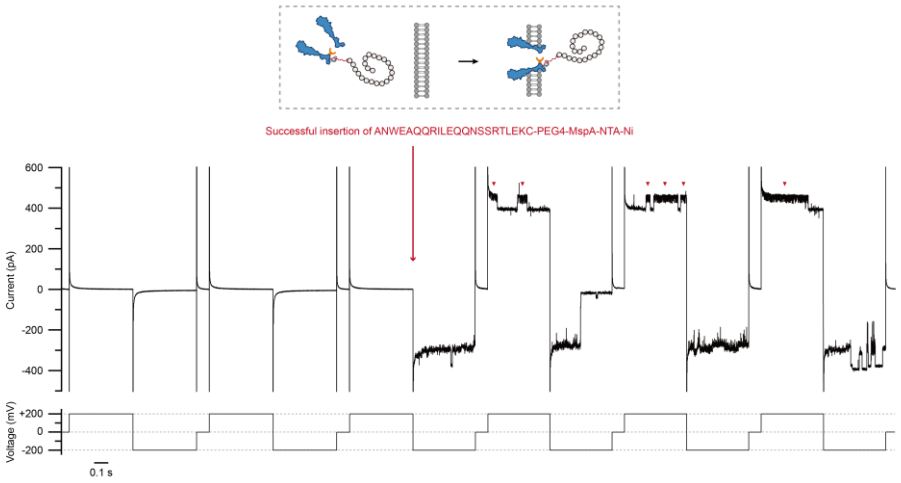


Figure RL 2.2. The spontaneous insertion of ANWEAQQRILEQQNSSRTLEKC-PEG4-MspA-NTA-Ni. A programmed cycling voltage protocol (+200 mV for 0.5 seconds, -200 mV for 0.5 seconds, 0 mV for 0.1 seconds) was employed to trigger the insertion of peptide-conjugated nanopore. Characteristic tPAL events were immediately observed upon the insertion of ANWEAQQRILEQQNSSRTLEKC-PEG4-MspA-NTA-Ni, confirming that a peptide conjugated pore can spontaneously insert.

Eventually, we again thank Reviewer 2 for raising this constructive comment. The above-mentioned result has been incorporated into the manuscript as **Supplementary Figure 12**. We believe that the inclusion of this result will fully address your concern regarding the insertion capacity of peptide attached nanopores.

Review Comments:

2. The nanopore signal is influenced by multiple amino acids, rendering sequencing a complex task. In their initial version, the authors implied that the signal is dominated by the last amino acid, implying the possibility of sequencing. Following my inquiry regarding homopolymeric peptides, the authors acknowledged that the signal is influenced by several amino acids, indicating that sequencing will be highly challenging. This issue is particularly relevant for small peptides, as natural peptides possess a vast sequence space. A recent study from BGI-Shenzhen, which employed a different but sequencing-compatible system (10.1038/s41467-026-69628-1), effectively illustrates this problem.

Response:

We sincerely appreciate reviewer 2 for raising the discussion regarding the spatial resolution of tPAL method. We wish to clarify that we never claimed that the signal was solely determined by the last amino acid. In the initial manuscript, we clearly stated that the event features are predominantly influenced by a few amino acids at the N-terminus (**Figure RL 2.3**). Following the constructive suggestions from all reviwers in the first round of review, we have conducted a series of experiments that further refined our understanding on the spatial resolution of tPAL method. The experimental results provide comprehensive evidence demonstrating that the tPAL signal is predominantly governed by the first three N-terminal amino acids, with decreasing contributions from subsequent residues (**Supplementary Figure 53-54**). Moreover, we have also developed a 3-mer-based decoding algorithm during the first revision, which provides direct evidence for the feasibility of sequence decoding using our platform.

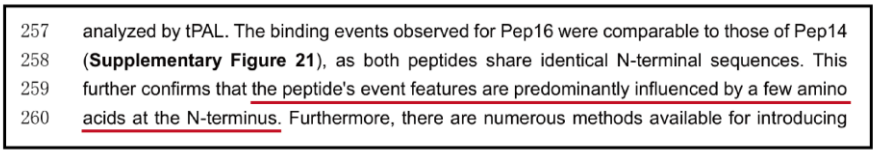

257 analyzed by tPAL. The binding events observed for Pep16 were comparable to those of Pep14
258 (**Supplementary Figure 21**), as both peptides share identical N-terminal sequences. This
259 further confirms that the peptide's event features are predominantly influenced by a few amino
260 acids at the N-terminus. Furthermore, there are numerous methods available for introducing

Figure RL 2.3. The description in our initial version of the manuscript.

The high spatial resolution of tPAL method originates from its novel design principle. Unlike conventional nanopore sequencing methods that rely on linear translocation through non-specific nanopore, tPAL method employs well-defined reversible chemical interactions that impose strict conformational constraints on the N-terminal region of the peptide, thereby ensuring that the peptide events are primarily contributed from a few N-terminal residues.

It is worth noting that the peptide analysis strategy used in the recent study of BGI was first reported by our group (Nano Letters, 2021, 21, 6703-6710). In the course of subsequent development, however, we identified fundamental limitations of this approach in achieving sufficient resolution for practical peptide sequencing, and our efforts have since shifted toward the development of reactive nanopores and its more advanced version: tPAL. As demonstrated in the present work, a reactive nanopore offers substantially enhanced resolution for peptide analysis and represents a more technically viable path forward.

Review Comments:

3. Long peptides exhibit a degree of structural organization, rendering the N-terminus less accessible. In my previous review, I requested the authors to provide data on the sequencing of a long peptide. Regrettably, they opted for an artificially constructed sequence: EQTREQTREQTREQTREQTREQTRC. This is disappointing, as the peptide lacks folding and comprises only four amino acids.

Most of these issues I have raised could have been resolved if the authors had characterized a long peptide with a random sequence. Notably, I previously requested in my review that the authors address the issue of secondary structure in long peptides. However, they chose to test EQTREQTREQTREQTREQTREQTRC, a non-natural peptide with a simplified amino acid composition and no secondary structure.

Consequently, despite their arguments, the authors have neglected to address the primary concern of this work. This is disappointing, as the manuscript currently presents a misleading portrayal, concealing the main issues with this approach.

To address this, I propose an experiment. The authors should attach a peptide of 20 or more amino acids derived from a lys-C digest of a natural protein (the author can select the peptide). The peptide can include an additional cysteine at the C-terminus to facilitate nanopore conjugation.

Then, they should demonstrate that a nanopore with such a peptide attached can insert into a membrane and that the entirety of the peptide can be read by the nanopore, distinguishing the twenty or more steps.

It is important to note that this is a simplified experiment that does not yet demonstrate sequencing. The C-terminal attachment is not addressed (the authors can use a cysteine at the C-terminus), and I am not requesting that the peptide be sequenced, but rather that each step can be distinguished.

If the authors can successfully perform this experiment, I would recommend that they submit the manuscript for publication.

Response:

We sincerely appreciate reviewer 2's insightful comments regarding long peptide detection. In the previous round of revision, we have provided stepwise cleavage data of the 25-mer peptide (EQTREQTREQTREQTREQTREQTRC) to address your concern about the detection capability of tPAL method toward longer peptides. The selection of this tandem repeat sequence was based on its ability to generate distinctive periodic event pattern during the sequential cleavage process, which unambiguously reports the progress of the peptide cleavage which serves to confirm the success of peptide cleavage all the way to the last amino acid.

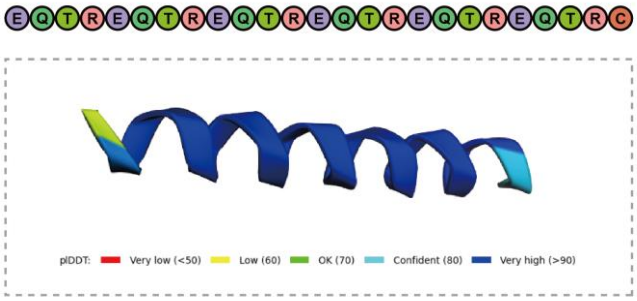
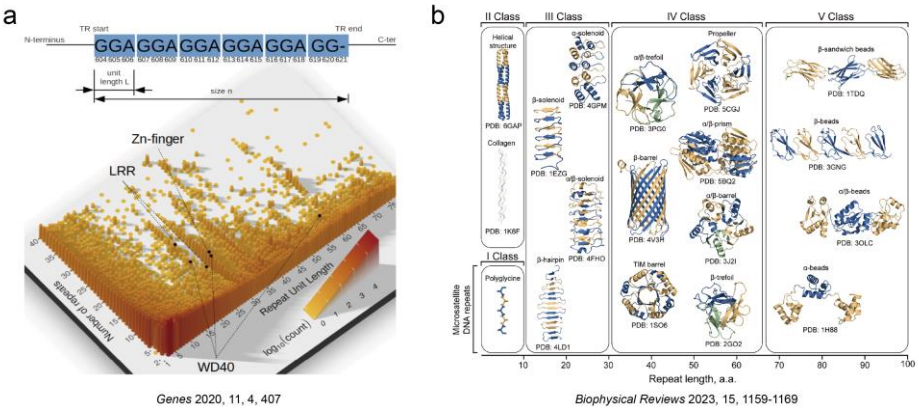


Figure RL 2.4. The 3D structure of EQTREQTREQTREQTREQTREQTRC predicted by Alphafold2. The prediction result demonstrates that this sequence adopts an α-helical conformation.

We sincerely apologize for failing to recognize the reviewer's inquiry that your concerns extended beyond read length to encompass peptide structure as well. However, although EQTREQTREQTREQTREQTREQTRC is not a random sequence, it does in fact possess a structure. Structural prediction via AlphaFold2 indicates that this peptide adopts an α-helical conformation and a high level of confidence score is reported in the prediction (**Figure RL 2.4**). It is also reported that tandem repeats are highly abundant in the human proteome (more than 60%) and display an impressive variability of sizes and structures (**Figure RL 2.5**). Therefore, we believe that the results of EQTREQTREQTREQTREQTREQTRC sequential cleavage have already demonstrated the tPAL method's ability to detect long peptides with structures.



Genes 2020, 11, 4, 407 *Biophysical Reviews* 2023, 15, 1159-1169

Figure RL 2.5. The pervasiveness and structure of tandem repeat sequences. (**a**) The distribution of all annotated tandem repeats, categorized by repeat unit length and number of repeats. (**b**) The structure of proteins containing tandem repeats of varying lengths.

However, to show respect of your inquiry, strictly following your request, we have performed additional experiments in this revision using a long peptide found in a natural protein. Specifically, the 22-mer peptide (ANWEAQQRILEQQNSSRTLEKC) was derived from the human septin-7 protein (residues 409-429), with an additional cysteine added at the C-terminus to enable peptide conjugation with the nanopore. During its tPAL measurement, we clearly observed 22 distinct event stages corresponding to the stepwise cleavage of the peptide from the N-terminus to the C-terminal cysteine residue (**Figure RL 2.6**). This result fully demonstrates the tPAL method's capability for detecting long peptides and highlights its promising potential for applications in protein analysis.

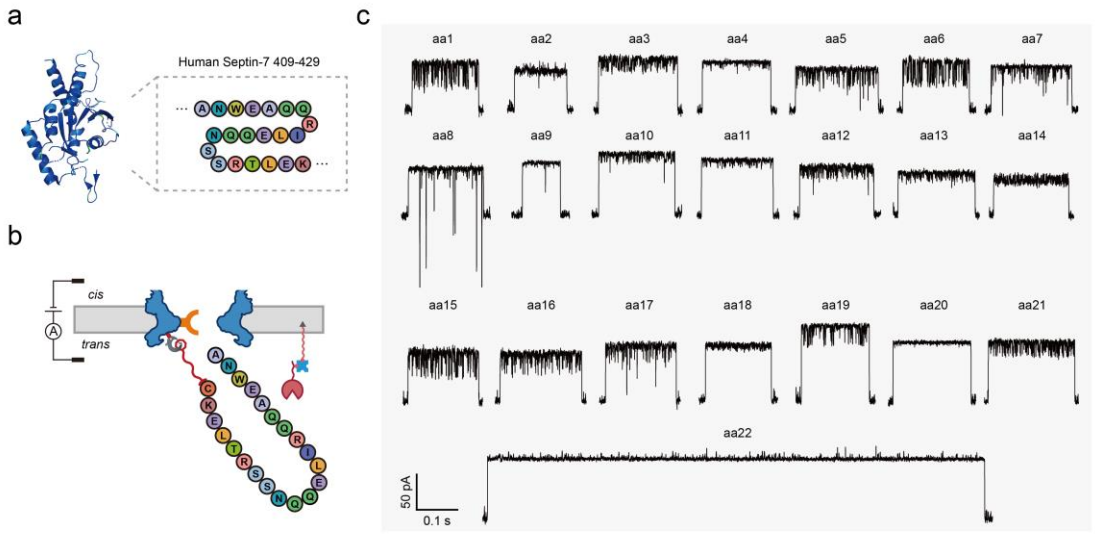


Figure RL 2.6. Sequential cleavage of ANWEAQQRILEQQNSSRTLEKC. (**a**) The model peptide with sequence of ANWEAQQRILEQQNSSRTLEK was selected from human septin-7 (409-429). (**b**) Schematic diagram depicting sequential cleavage of ANWEAQQRILEQQNSSRTLEKC using sgAP-dA15-chol. (**c**) Representative events acquired with ANWEAQQRILEQQNSSRTLEKC during sequential cleavage. Stepwise cleavage of the peptide from the N-terminus to the C-terminal cysteine residue was fully observed, resulting in 22 stages labeled as aa1-aa22, respectively.

We sincerely appreciate Reviewer 2 for raising this constructive discussion, which has significantly enhancing the scientific depth and rigor of our method. To fully address this issue, relevant result and discussion have been incorporated into the revised manuscript and are pasted below for the ease of your reference:

*"A read length of up to 22 amino acids was achieved (**Supplementary Figure 12, 59**), with longer reads also deemed feasible but not experimentally examined herein."*

Finally, we sincerely thank Reviewer 2 for the time and effort devoted to multiple rounds of review. We are eager to share this methodology with the broader community, with the hope of inspiring collaborative efforts that will collectively advance the technology, such as sequencing throughput, natural peptide conjugation, and sequence decoding algorithms. We also wish to note that the first author has devoted nearly six years to this work with exceptional dedication, and her graduation has already been delayed. Your support to the

publication of this work is sincerely acknowledged.

Response Letter

Reviewer #2:

Review Comments:

The authors have conducted the experiment I requested, selecting the 22-mer peptide ANWEAQQRILEQQNSSRTLEKC from human septin-7. The peptide contains an aromatic residue (tryptophan at position 3) alongside isoleucine and two leucines, which support a modest hydrophobic core and some degree of helical structure—consistent with its predicted α-helical conformation in the intact septin-7 protein. To more rigorously address the folded issue, a peptide enriched in multiple aromatic (Trp, Phe, Tyr) and aliphatic hydrophobic residues (Leu, Ile, Val, Pro) would have provided a more stringent test of the system's ability to handle structured substrates, where N-terminal accessibility to the nanopore and the aminopeptidase may be genuinely compromised. Nevertheless, in light of the data presented, I think this peptide provides a reasonable demonstration that a peptide with a scrambled sequence can be addressed by the system.

Response:

We are deeply grateful to the reviewer for the insightful and constructive suggestion regarding the application of a naturally occurring peptide to demonstrate the system's analytical capabilities. The experimental data obtained through this approach have significantly enhanced the scientific rigor and overall quality of our work. We believe that this result could well demonstrate the generality, read length and resolution of this technique and your suggestion is highly acknowledged.

We are also particularly appreciative that you have confirmed the validity of this newly performed measurements and have acknowledged these results as a reasonable demonstration of the system's capacity to process peptides with scrambled sequences, as highlighted in your commentary. Your meticulous review and the substantial time invested in evaluating our manuscript are sincerely appreciated.

Review Comments:

The authors do not, however, provide a detailed interpretation of the ionic current traces in the rebuttal, which is frustrating. My interpretation by inspection of the traces in the figure provided reveals that the ionic current signatures for two chemically identical glutamine pairs differ significantly from one another. If true, this clearly indicates that the nanopore does not read a single terminal amino acid in isolation; rather the ionic current at each cleavage stage encodes the combined contribution of several adjacent residues —a k-mer effect. On one hand, this findings suggests the system is capable of reading short homopolymeric sequences. On the other hand, it reveals that the signal reflect several

amino acids simultaneously, which complicates the system's ability to sequence generic peptides—a point I return to below.

An experiment with a scrambled amino acid sequence is therefore important and should be presented in the main text with figures, as it would provide a clear picture of the capabilities and current limitations of this methodology.

As it stands, in the section "Direct decoding of peptide sequence," the authors do acknowledge the existence of a k-mer effect:
1) “Based on the insight that peptide events are predominantly influenced by the first few amino acids at the N-terminus” (lines 348-349).
2) “…suggesting that the first three N-terminal amino acids play a dominant role in determining event features” (lines 344-345).
3) “Notably, the tPAL events in each stage encode information not only about the N-terminal amino acid of that stage but also about few downstream residues” (lines 361-362).

However, on a first reading, these statements are unclear, as no data are shown for generic peptides—only a cryptic and superficial Figure S54. The new result with the natural-sequence peptide—currently buried in the SI with minimal discussion—makes the meaning of these statements clear, and also explains why all sequencing demonstrations in this manuscript employed repeat-sequence peptides. The authors have been validating their system precisely under conditions where its most fundamental limitation is invisible. This is an important finding that the community deserves to understand fully, and it should not be concealed in a supplementary figure with minimal discussion.

Therefore, the authors should present in a main-text figure an experiment—which could be the new data recently collected—showing the signal from a scrambled peptide sequence, described in detail. They should explicitly state that the same amino acid is read differently depending on its sequence context. This is a key feature that the reader must grasp while reading the article. The authors should then openly discuss whether protein sequencing can be achieved with this approach, and how. This should not be buried in an eloquent but vague passage, nor relegated to a supplementary figure.

Response:

We thank Reviewer 2 for the constructive suggestion to present the newly added results as a main-text figure. We would like to note that in the previous round of revision, this result was placed in the SI in response to the editorial team’s request to streamline the manuscript and reduce the number of maintext display items at that time. Nonetheless, to fully address the reviewer’s specific request, we have made the following revisions:

1. We have moved the previous maintext Figure 4 to the Extended Data to accommodate the mentioned figure result.

2. The SI figure referenced by the reviewer has been moved to the main text as the new **Figure 6** in the maintext
3. On top of the raw ionic current trace, we have clearly annotated the corresponding k-mer sequence in a step-by-step manner above each event (**Figure RL.1**).
4. We have added a corresponding figure legend and relevant discussion in the main text to clearly describe the corresponding k-mers.
5. We explicitly describe that the same amino acid is read differently due to its sequence context. This description is clearly provided in the corresponding figure legend:

"Notably, subtle differences in event features can be observed for the same terminal amino acid (such as QQR and QRI), owing to the influence of its sequence context, a property that proves advantageous for resolving short homopolymers (Extended Data Figure 9)"

Notably, as also mentioned by the reviewer, this capacity is technically advantageous for resolving short homopolymers as well and in the last round of revision, we have already provided corresponding experimental results and corresponding discussions for this topic. In this revision, it is demonstrated as the new **Extended Data Figure 9** as well.

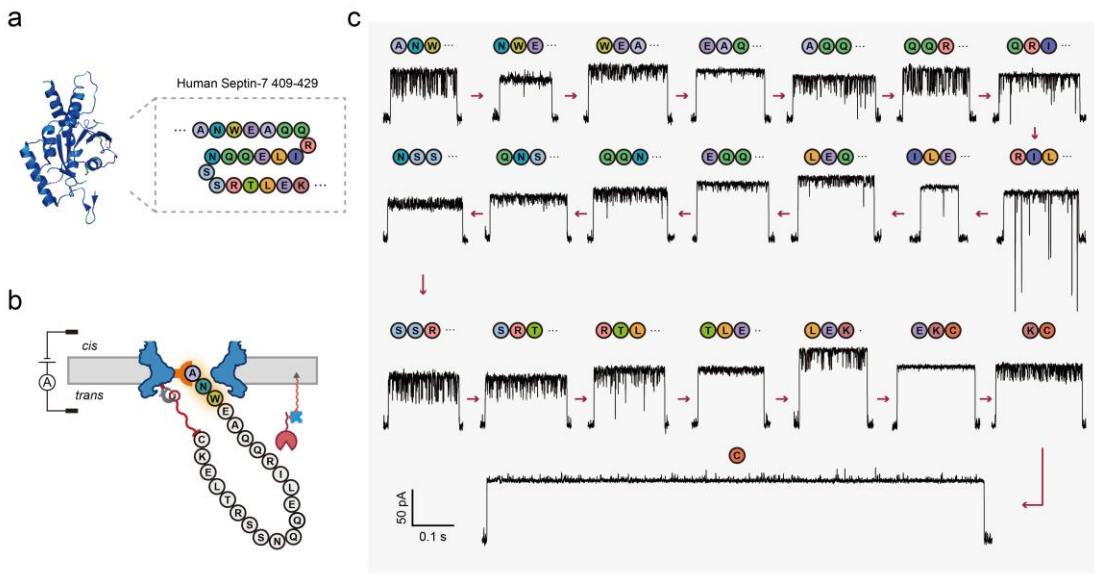


Figure RL.1 Sequential cleavage of long peptide. (**a**) The peptide with sequence of ANWEAQQRILEQQNSSRTLEK was selected from human septin-7 (residues 409-429). (**b**) Schematic diagram depicting sequential cleavage of ANWEAQQRILEQQNSSRTLEKC using sgAP-dA15-chol. The tPAL events generated at each stage primarily encode information about the first three amino acids. (**c**) Representative events acquired with ANWEAQQRILEQQNSSRTLEKC during sequential cleavage. Stepwise cleavage of the peptide from the N-terminus to the C-terminal cysteine residue was fully observed, resulting in 22 stages. The corresponding k-mer read at each stage is labeled above each event. Notably, subtle differences in event features can be observed for the same terminal amino acid (such as QQR and QRI), owing to the influence of its sequence context, a property that proves advantageous for resolving short homopolymers (**Extended Data Figure 9**).

We would also like to take this opportunity to reiterate that, ever since the initial submission of this manuscript, we have explicitly stated that tPAL events are governed by the first few amino acids acting as a k-mer, with certain amino acids (E, F, H, P, Q, etc.) at the N-terminal position playing a critical role in event feature generation so that their presence could be easily recognized. In the previous revision round, we included dedicated main-text figures (renumbered as **Figure 5** in the current version) and discussion on peptide sequence decoding and the role of the corresponding k-mers, which was also acknowledged in the reviewer's comments (lines 348-349, 344-345, and 361-362), as well as in **Supplementary Figures 52-54**.

We would further like to highlight that our tPAL analysis demonstration was not limited to repeat-sequence peptides. As shown in **Figure 5**, peptides such as ETREFTRSC and KTREFTRSC contain no sequence repetition, and a 3-mer model successfully decoded their signals with high accuracy, achieving precise interpretation from raw ionic current to amino acid sequence. Some discussion on the potential of tPAL method for protein sequencing are also provided in the conclusion section, underscoring its applicability beyond synthetic peptides and providing a complete and unambiguous response to the reviewer's comments. For the ease of reference, they are pasted below for the ease of your reference:

"The capability of tPAL to resolve long peptides of random sequence was validated with ANWEAQQRILEQQNSSRTLEKC. This peptide sequence is derived from an enzymatically cleaved fragment of human septin-7 with an appended C-terminal cysteine, mimicking the tPAL analysis of native proteins after proteolytic hydrolysis. As shown in ***Figure 6****, 22 distinct event stages, each encoding the sequence information of multiple corresponding N-terminal residues rather than a single N-terminal amino acid, were clearly identified during sequential enzymatic cleavage. This demonstrates the successful tracking of prolonged consecutive cleavage steps and validates the method's potential for protein sequencing."*

"Though demonstrated here with synthetic peptides, tPAL method holds great promise for the analysis of natural proteins. Proteins can be first hydrolyzed into peptide fragments using endopeptidases and then immobilized to a nanopore through photocatalytic C-terminal-specific labeling methods[46] to enable tPAL measurements."

Review Comments:

The issue with K-mer in peptides.
If, as superficially mentioned, only three amino acids contribute to each signal—which is already an optimistic estimate, given that the k-mer for DNA nanopore sequencing is typically 5 and estimates for peptide k-mers place it around 8—then 20^3 = 8,000 distinct signal states must be deconvolved to sequence an arbitrary peptide from each single read. This makes sequencing highly challenging, given the vast combinatorial space of natural peptide sequences. If the effective k-mer is larger, sequencing may not be feasible at all. Nevertheless, as demonstrated in DNA nanopore sequencing, the step-wise nature of

enzymatic cleavage simplifies the overall decoding problem. Furthermore, repeated reading of the same (cleavage) position—enabled here by multiple binding/unbinding events at each stage—should reduce stochastic error in k-mer calling.

Response:

We sincerely appreciate Reviewer 2's insightful discussion regarding the challenges of k-mer space and sequence decoding.

Regarding your comment:

"Nevertheless, as demonstrated in DNA nanopore sequencing, the step-wise nature of enzymatic cleavage simplifies the overall decoding problem. Furthermore, repeated reading of the same (cleavage) position—enabled here by multiple binding/unbinding events at each stage—should reduce stochastic error in k-mer calling."

The context of this sentence is best understood in the framework of peptide nanopore sequencing (demonstrated in this study using tPAL) rather than DNA nanopore sequencing. This clarification arises from two key technical distinctions:

1. Current DNA nanopore sequencing relies on helicase-mediated strand unwinding instead of enzymatic cleavage, which is unique and central to this tPAL technique.

2. The repeated reading of the same cleavage site through multiple binding/unbinding events, a key feature enabling improved sequence resolution, is uniquely enabled by the tPAL methodology and not achievable with any existing DNA nanopore sequencing technologies.

While the reference to DNA sequencing is likely a typographical error, we fully acknowledge the reviewer's intent to highlight the technical advantages of the tPAL technique, specifically its capacity to simplify the overall decoding architecture and mitigate stochastic errors during sequence decoding. In response to this constructive feedback, we have systematically incorporated these advantages into our revised manuscript through a comprehensive discussion of the methodological innovations in the technical framework section. For the ease of your reference, it is pasted below:

"This constitutes a major advantage of tPAL, achieved through enzymatic cleavage and repeated N-terminus readout of cleaved peptides, which simultaneously simplifies the decoding architecture and mitigates stochastic errors in k-mer calling."

Regarding the k-mer space size and challenges in sequence decoding, recent advancements in nanopore nucleic acid sequencing indicate a predominant trend toward selecting k values of 6-9 (rather than k = 5) for k-mer analysis (**Figure RL.2,** *Nat Methods* 22, 681-691 (2025)). The corresponding k-mer space ranges between 4^6=4096 and

4^9=262144, comparable or significantly larger than the 20^3=8000 distinct k-mers required for peptide sequence decoding as suggested in this manuscript.

Figure RL.2 The literature evidence of k-mer selection in mainstream ONT chips. According to technical details disclosed in relevant literatures, the r9.4 and r10.4 chips offered by oxford nanopore technology (ONT) respectively utilizes 6-mer and 9-mer for sequence decoding (*Nat Methods* 22, 681-691 (2025)).

We speculate that the adoption of large k-mers in mainstream nanopore nucleic acid sequencing reflects the inherent limitations of current pore resolution and the lack of well-defined pore-analyte chemical interactions. The resulted stochastic errors and sequence-context interference necessitate a larger k-mer space to achieve reliable decoding accuracy. This also reflects the capacity of modern computational algorithms to accommodate and effectively process such large k-mer models. In different words, this suggests that the k-mer space for peptide sequencing is not inherently more complex than the challenges already addressed in contemporary nanopore nucleic acid sequence decoding algorithms and the technical advantage of tPAL in requiring even smaller k-mer spaces than the current mainstream nanopore nucleic acid sequence decoding algorithms.

Notably, we also observed that specific N-terminal residues (e.g., E, R, H, and P) at the protein sequence's N-terminus generate signals with conserved characteristics. This phenomenon is likely due to their unique interactions with Ni^{2+} and the pore lumen environment. These consistent signal patterns could substantially simplify decoding by reducing ambiguity in sequence inference. Such improvements may also enhance decoding accuracy and diminish the reliance on extensive, computationally intensive training datasets, thereby advancing the efficiency of nanopore-based peptide sequencing.

Finally, we again acknowledge Reviewer 2 for raising this extremely inspiring discussion and your appraise of the technical advantage of tPAL for k-mer calling simplification and stochastic error reduction. To fully address your comment, we have incorporated extended discussions of k-mer space along with your suggested discussions in the revised manuscript. For your convenience, we have included the relevant section below:

*"This suggests direct peptide sequence decoding from tPAL results through a 3-mer-based decoding strategy (**Figure 5a**). This approach is conceptually similar to the basecalling methodology used in nanopore nucleic acid sequencing, where k-values of 6-9 are commonly*

used to achieve reliable nucleic acid sequence identification, corresponding to a reference database covering 4^6 = 4,096 to 4^9 = 262,144 combinations. In contrast, the high spatial resolution achieved by tPAL enables peptide sequence decoding at 3-mer level. The corresponding database for all possible 3-mer combinations contains 20^3 = 8,000 entries, readily achievable using modern bioinformatics pipelines through efficient pattern recognition algorithms and machine learning frameworks."