Supplementary Materials for

Multiplexed direct detection of barcoded protein reporters on a nanopore array

Nicolas Cardozo, Karen Zhang, Kathryn Doroschak, Aerilynn Nguyen, Zoheb Siddiqui, Nicholas Bogard, Karin Strauss, Luis Ceze and Jeff Nivala

Correspondence to: jmdn@uw.edu

This PDF file includes:

Figs. S1 to S13
Supplementary Notes

NTER amino acid sequence (*E. coli*):

MTMTRLKISKTLLAVMLTSAVATGSAYAENNAQTTNESAGQKVDSSMNKVGNFMDDSAITAKVKAALVDHDNIKSTDISVKTDQKVVT
LSGFVESQAQAEEAVKVAKGVEGVTSVSDKLHVRDAKEGSVKGYAGDTATTSEIKAKLLADDIVPSRHVKVETTDGVVQLSGTVDSQ
AQSDRAESIAKAVDGVKSVKNDLKTK MGHHHHHHHHHHHGS LQDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRR
LMEAFAKRQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQI XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX DGGSSGG
SGGDGSSGDGGSDGDSDGSDGDGDSDGDDGGDDEDDGSDD*

mNTER amino acid sequence (*HEK293*):

MALTFALLVALLVLSCKSSCSV GGHHHHHHHHHH VSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKV
TKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPV
MQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHS
TGGMDELYKSGS LQDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLTFLYDGIEIQADQ
APEDLDMEDNDIIEAHREQI XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX DGGSSGGSGGDGSSGDGGSDGDSDGSDGDGDSD
GDDGGDDEDDGSDD*

```
NTERY00 X = GGGGSSGGSGGSGSSGDGGSSGGSGGSGSSG
```

YYY mapping mutants:

```
NTERY01 X = YYYGSSGGSGGSGSSGDGGSSGGSGGSGSSG
NTERY02 X = GGYYYSGGSGGSGSSGDGGSSGGSGGSGSSG
NTERY03 X = GGGGYYYGSGGSGSSGDGGSSGGSGGSGSSG
NTERY04 X = GGGGSSYYYGGSGSSGDGGSSGGSGGSGSSG
NTERY05 X = GGGGSSGGYYYSGSSGDGGSSGGSGGSGSSG
NTERY06 X = GGGGSSGGSGYYYSSGDGGSSGGSGGSGSSG
NTERY07 X = GGGGSSGGSGGSYYYGDGGSSGGSGGSGSSG
NTERY08 X = GGGGSSGGSGGSGSYYDGGSSGGSGGSGSSG
NTERY09 X = GGGGSSGGSGGSGSSGDYYSSGGSGGSGSSG
NTERY10 X = GGGGSSGGSGGSGSSGDGYYYGGSGGSGSSG
NTERY11 X = GGGGSSGGSGGSGSSGDGGSYYYSGGSGSSG
NTERY12 X = GGGGSSGGSGGSGSSGDGGSSGYYYGSGSSG
NTERY13 X = GGGGSSGGSGGSGSSGDGGSSGGSYYYGSSG
NTERY14 X = GGGGSSGGSGGSGSSGDGGSSGGSGGYYYSG
NTERY15 X = GGGGSSGGSGGSGSSGDGGSSGGSGGSGYYY
```

Homopolymer mutants:

```
NTER A X = GGAAAAAAAAASGSSGDGGSSGGSGGSGSSG
NTER D X = GGDDDDDDDDDSGSSGDGGSSGGSGGSGSSG
NTER E X = GGEEEEEEEEESGSSGDGGSSGGSGGSGSSG
NTER G X = GGGGGGGGGGGSGSSGDGGSSGGSGGSGSSG
NTER H X = GGHHHHHHHHHSGSSGDGGSSGGSGGSGSSG
NTER M X = GGMMMMMMMMSGSSGDGGSSGGSGGSGSSG
NTER N X = GGNNNNNNNNNSGSSGDGGSSGGSGGSGSSG
NTER P X = GGPPPPPPPPPSGSSGDGGSSGGSGGSGSSG
NTER Q X = GGQQQQQQQQQSGSSGDGGSSGGSGGSGSSG
NTER R X = GGRRRRRRRRRSGSSGDGGSSGGSGGSGSSG
NTER S X = GGSSSSSSSSSSGSSGDGGSSGGSGGSGSSG
NTER T X = GGTTTTTTTTTSGSSGDGGSSGGSGGSGSSG
```

PKA motif mutants:

```
NTER PKA                 X = GGRRGSYYSGGSGSSGDGGSSGGSGGSGSSG
NTER PKA_phosphomimetic X = GGRRGEYYSGGSGSSGDGGSSGGSGGSGSSG
```

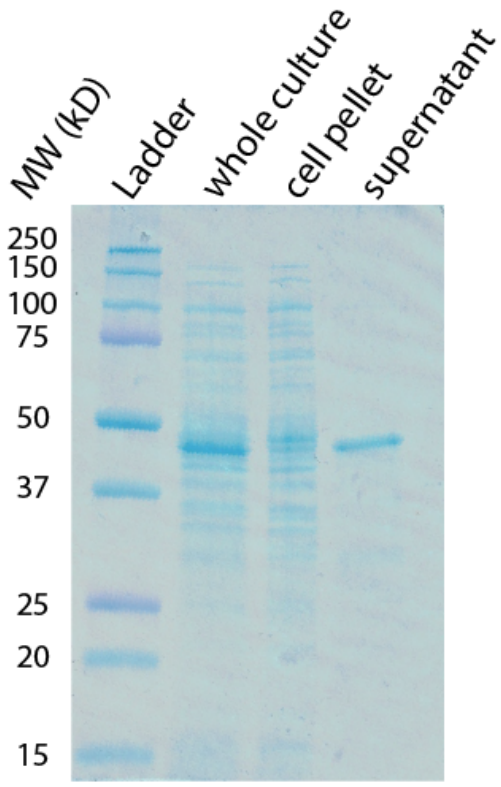
Tail mutants:

```
NTER truncated 40 tail = GDGSSGDGGSDGDSDGSDGDGDSDGDDGGDDEDDGSDD* (-18 charge)
NTER truncated 20 tail = GGGGSSGGSGGSGSSGDGGSSDGSSGDGGSDGDSDGSDGDGDSDGDDGGDDEDDGSDD*
(-19 charge)
NTERY00 standard tail =
GGGGSSGGSGGSGSSGDGGSSGGSGGSGSSGDGGSSGGSGGDGSSGDGGSDGDSDGSDGDGDSDGDDGGDDEDDGSDD* (-20
charge)
```

Supplementary Figure 1 NanoporeTER sequences used in this work. Blue: Secretion domain. Black: His-tag. Green: Smt3 domain (Smt3* in mNTER). Red: Variable region (contains barcode). Orange: PolyGSD tail. Purple: mCherry domain (only present in mNTER for convenience).

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTT
CTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACT
TGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAAC
AACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAA
CAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTA
ATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCAC
CGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTG
AGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCG
CATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACA
CTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTAC
GGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACA
ACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTA
ATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGT
GAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTA
CCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTA
GGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCA
AGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAG
CTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGG
TATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTA
CGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC
AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCA
CTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGC
GCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCA
GCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGAT
AAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCT
CACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATG
CCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACT
TTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACC
AGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTT
GGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAAT
GACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTT
GAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCG
CCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTAT
CGTCGTATCCCACTACCGAGATGTCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAA
CGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCC
AGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTT
CATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCG
GATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCA
GTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTG
CCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAG
AGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCC
ATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCA
TGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTG
ATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCGATCTCGATCCCGCGAAAT**TAATACGA**
CTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGACTATGACAAGACTGAAGATTTCGAAAA
CTCTGCTGGCTGTAATGTTGACCTCTGCCGTCGCGACCGGCTCTGCCTACGCGGAAAACAACGCGCAGACTACCAATGAAAGCGCAGGGCAAAAAGTCGATAGCTCTATGA
ATAAAGTCGGTAATTTCATGGATGACAGCGCCATCACCGCGAAAGTGAAGGCGGCCCTGGTGGATCATGACAACATCAAGAGCACCGATATCTCTGTAAAAACCGATCAAA
AAGTCGTGACCCTGAGCGGTTTCGTTGAAAGCCAGGCCCAGGCCGAAGAGGCAGTGAAAGTGGCGAAAGGCGTTGAAGGGGTGACCTCTGTCAGCGACAAACTGCACGTTC
GCGACGCTAAAGAAGGCTCGGTGAAGGGCTACGCGGGTGACACCGCCACCACCAGTGAAATCAAAGCCAAACTGCTGGCGGACGATATCGTCCCTTCCCGTCATGTGAAAG
TTGAAACCACCGACGGCGTGGTTCAGCTCTCCGGTACCGTCGATTCTCAGGCACAAAGTGACCGTGCTGAAAGTATCGCCAAAGCGGTAGATGGTGTGAAAAGCGTTAAAA
ATGATCTGAAAACTAAGATGGGTCACCACCACCACCACCACCACCACCACCACGGTAGCCTGCAAGATTCGGAAGTCAACCAAGAAGCAAAGCCGGAAGTCAAGCCGGAAG
TGAAACCGGAAACCCATATTAACCTGAAAGTTAGTGACGGCAGCTCTGAAATTTTCTTTAAGATCAAAAAGACCACGCCGCTGCGTCGCCTGATGGAAGCGTTTGCCAAAC
GTCAGGGCAAGGAAATGGATAGCCTGCGTTTCCTGTATGACGGTATTCGCATCCAGGCAGATCAAGCGCCGGAAGACCTGGACATGGAAGACAACGACATCATTGAAGCCC
ACCGTGAACAGATTGGTGGTGGTGGTTCGAGCGGCGGGAGCGGAGGGTCTGGGTCTAGCGGGGATGGTGGGTCGTCTGGGGGGAGCGGCGGTTCGGGTTCATCTGGTGACG
GCGGTTCCTCGGGTGGTTCCGGGGGAGATGGTTCTTCCGGAGATGGCGGGAGTGATGGGGACTCTGATGGTTCGGATGGTGACGGGGACAGTGATGGTGATGACGGAGGGG
ACGACGAGGATGACGGTTCAGACGATTAATGAGGCGGCTCGAGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAA*ACTAGC*
*ATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG*CTGAAAGGAGGAACTATATCCGGAT

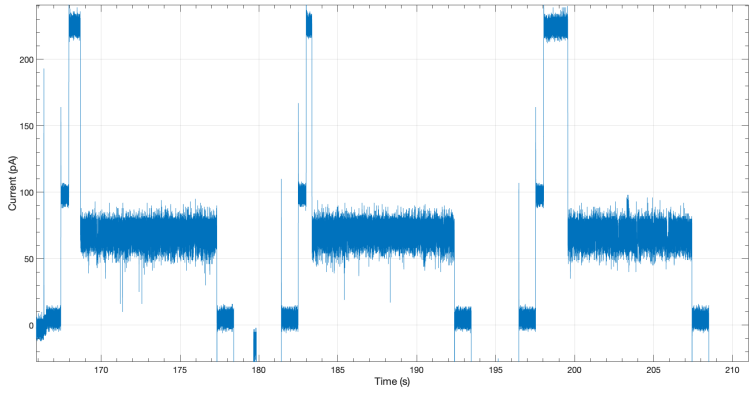
Supplementary Figure 1 (cont.) NanoporeTER (Y00) expression vector DNA sequence (based on the pCDB180 plasmid backbone). Annotations: T7 promoter and lacO (bold), RBS (underlined), NanoporeTER ORF (box), and T7 terminator (italic).



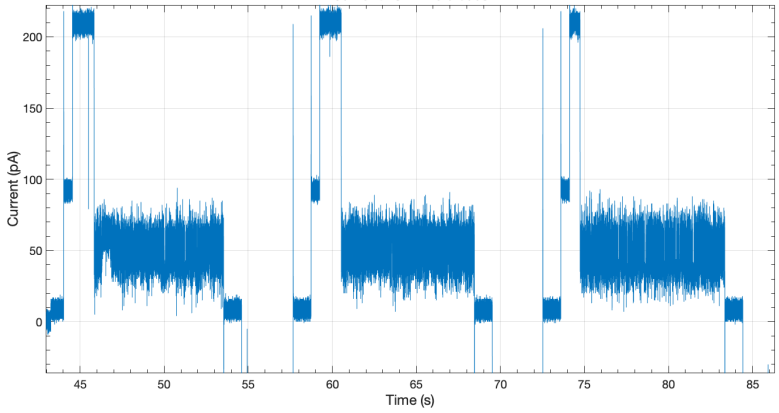
Supplementary Figure 2 NanoporeTERs are secreted into the extracellular medium. SDS-PAGE analysis of overnight culture of an E. coli strain transformed with a plasmid expressing NTER00 (expected MW is 40.2 kilodaltons). Lanes: 1, Ladder. 2, raw whole culture (cells and growth medium). 3, cell pellet resuspended in water following centrifugation. 4, Growth medium supernatant following centrifugation. Gel is representative of similar results for n=3 experiments.

Supplementary Figure 3 Representative MinION nanopore ionic traces of NTER barcodes analyzed in this work. Three consecutive "reads" are included in each trace.

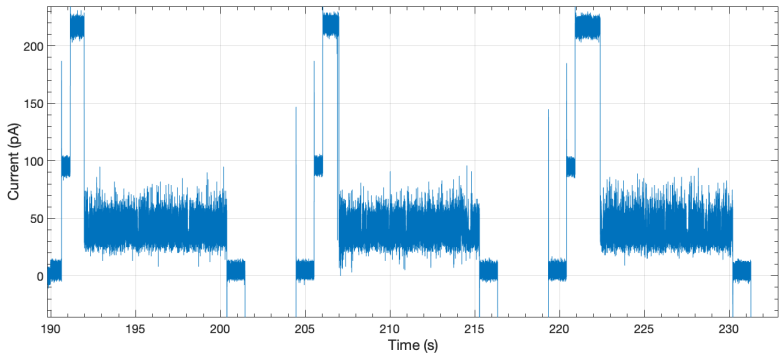
NTER00



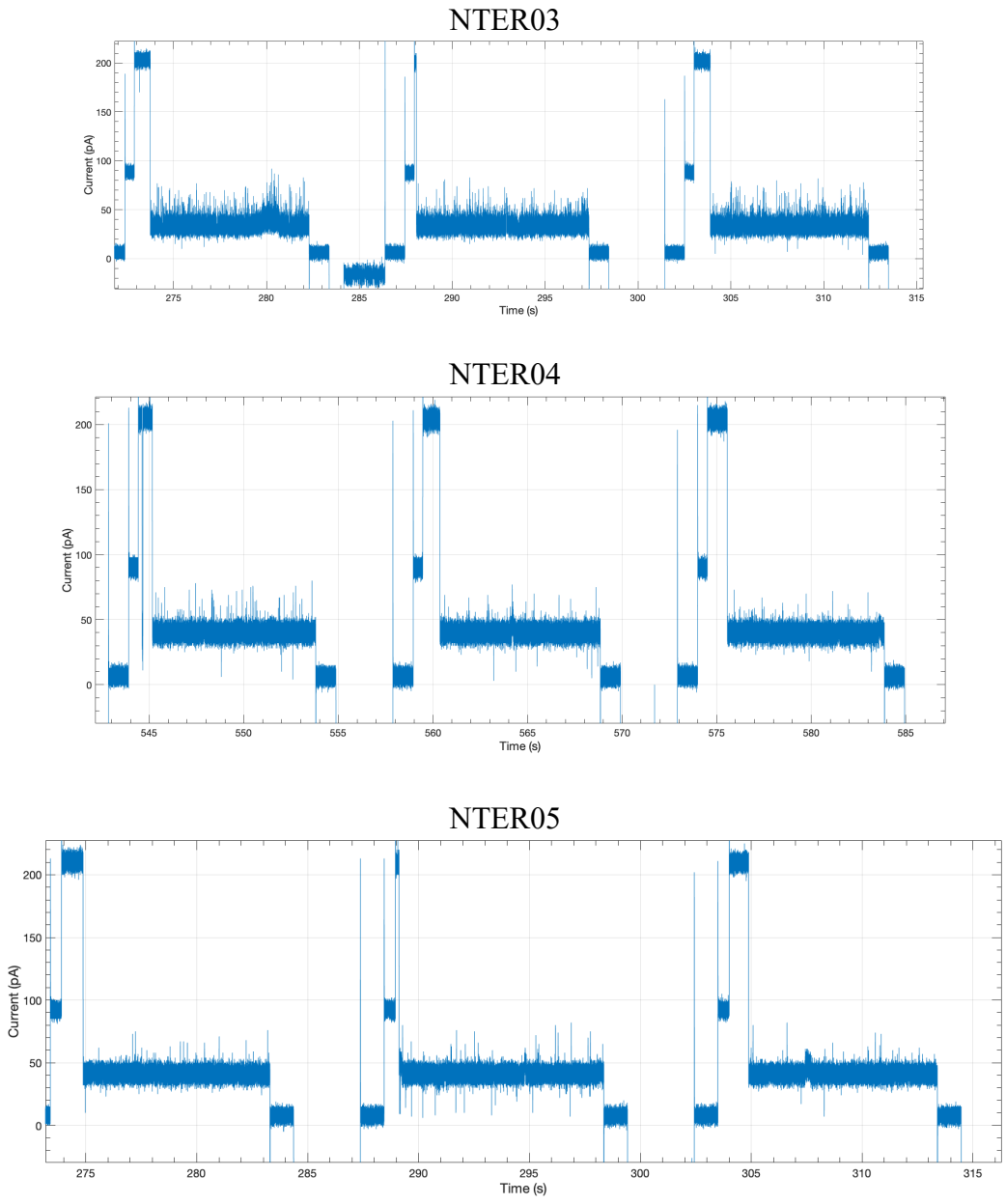
NTER01



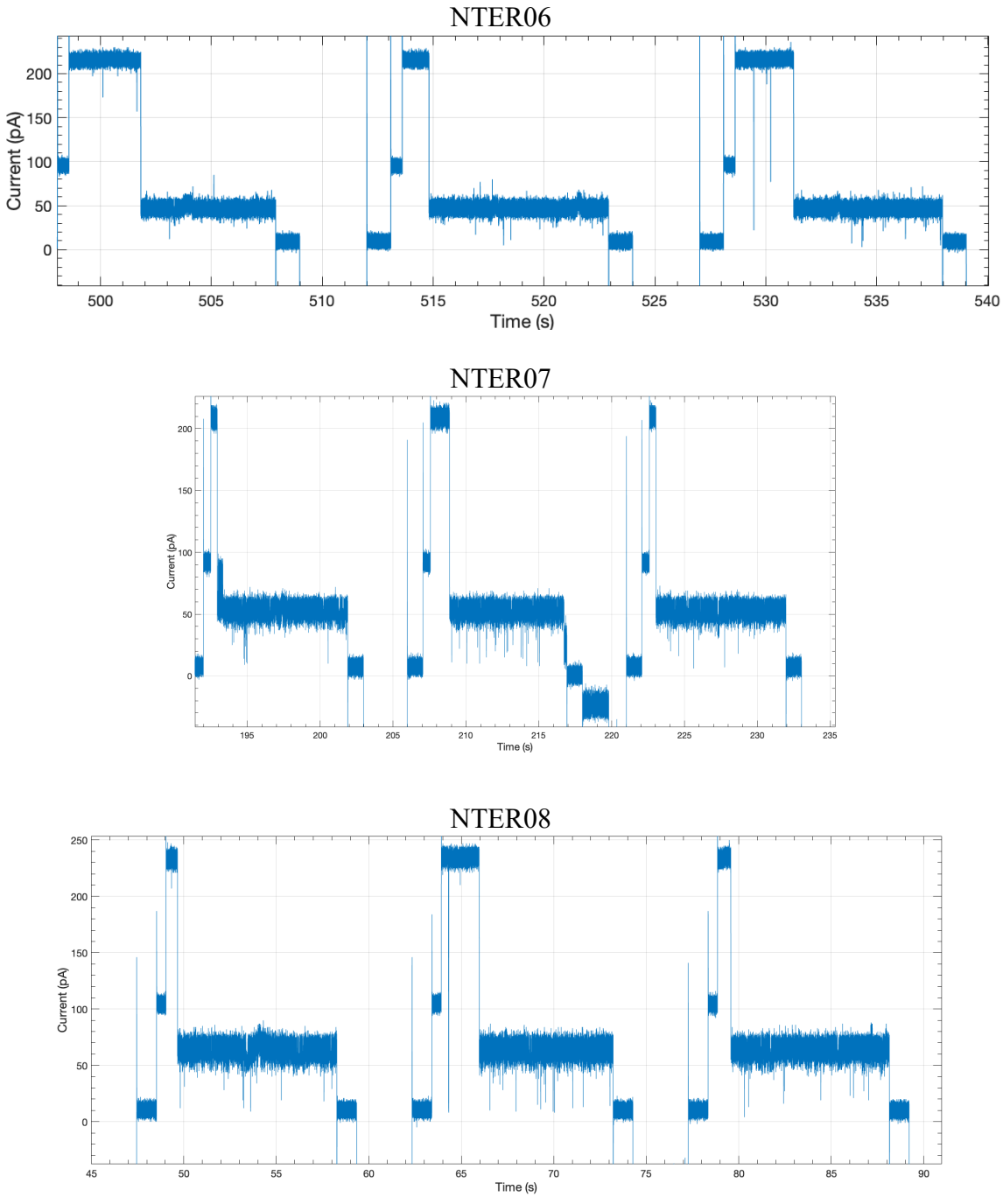
NTER02



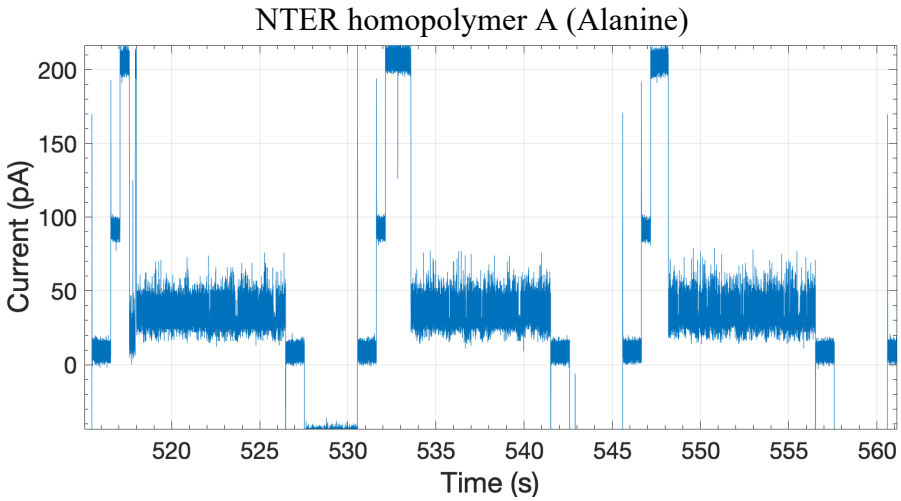
Supplementary Figure 3 (cont.)

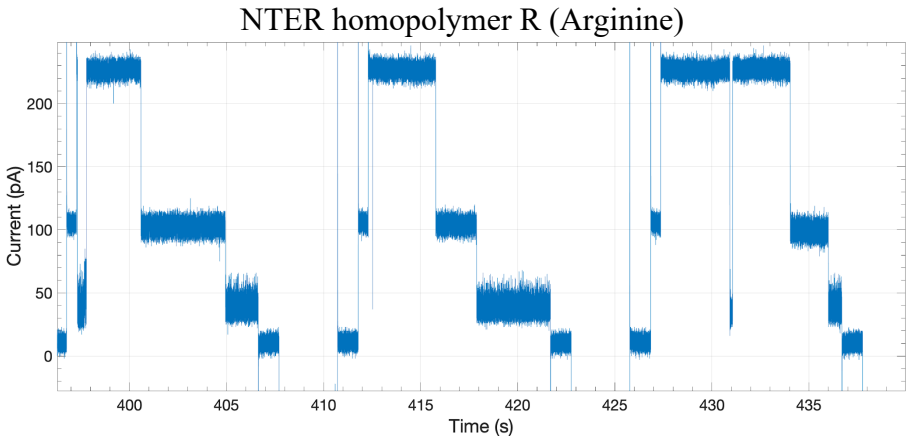


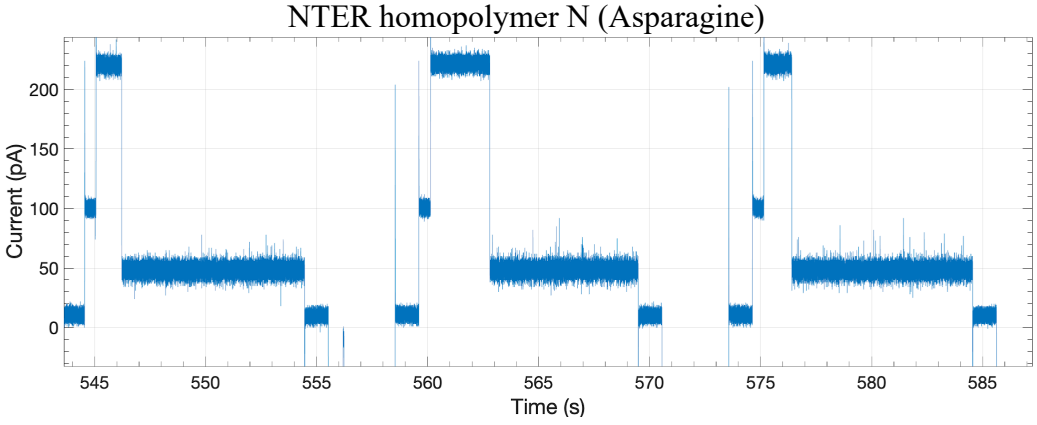
Supplementary Figure 3 (cont.)



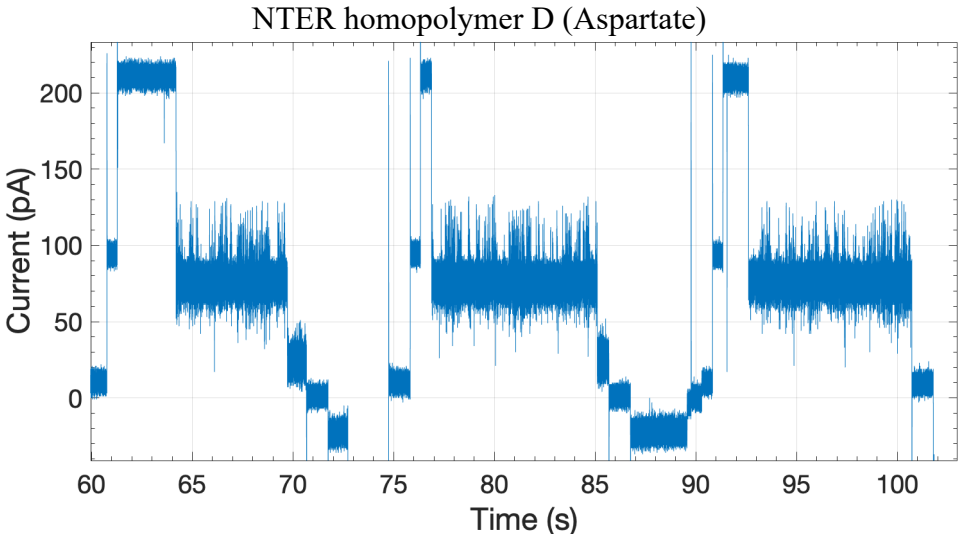
Supplementary Figure 3 (cont.)

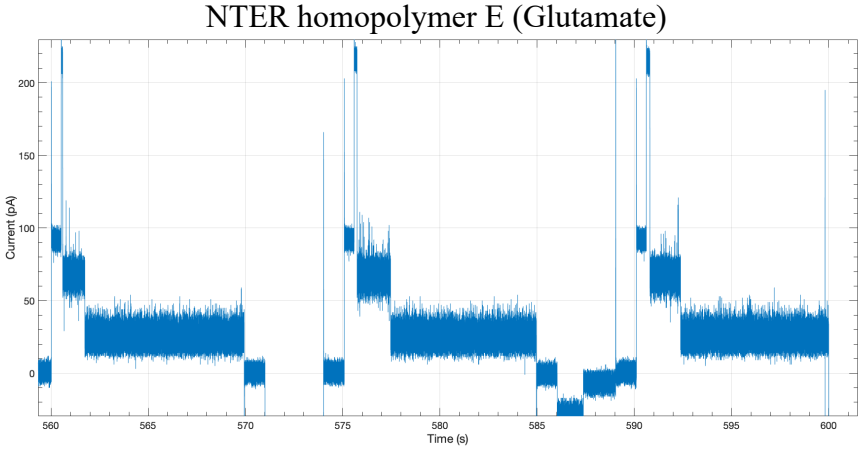


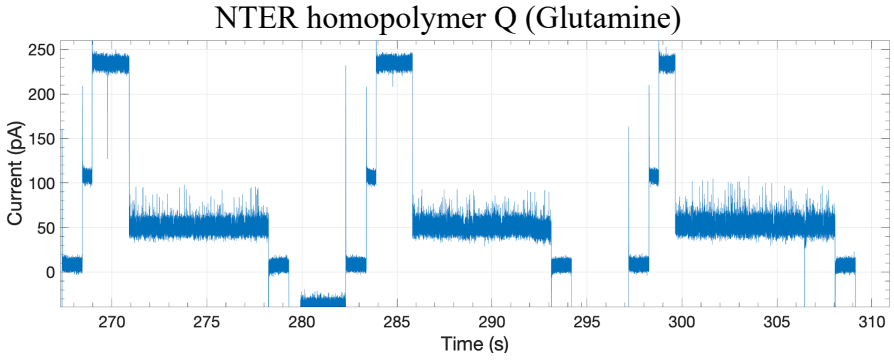




Supplementary Figure 3 (cont.)

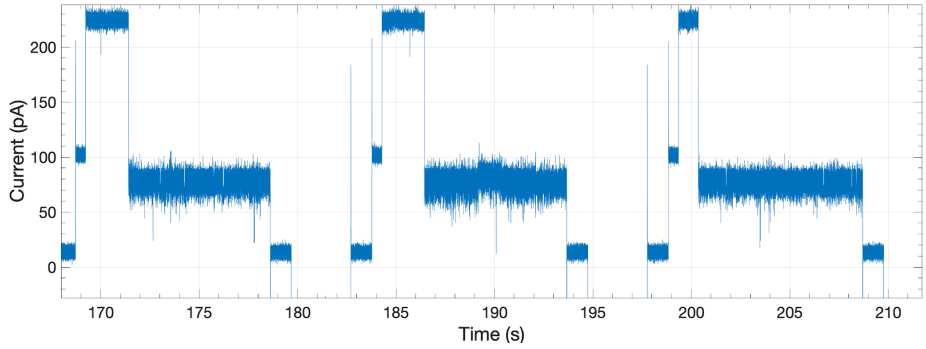




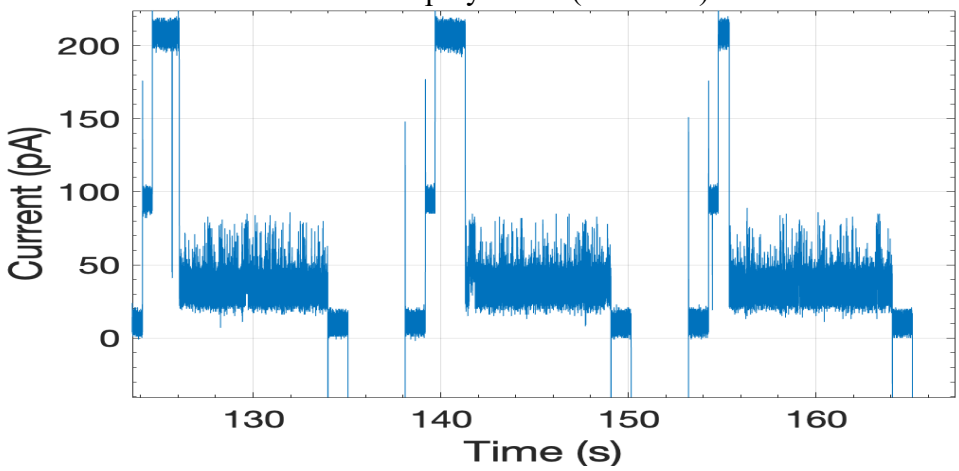


Supplementary Figure 3 (cont.)

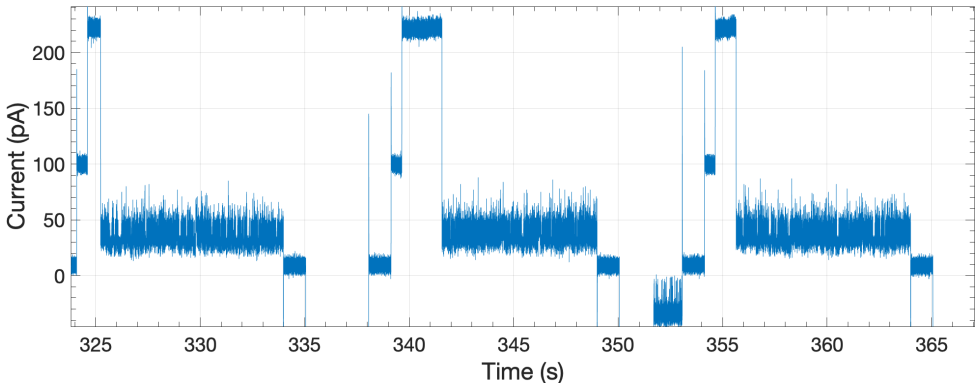
NTER homopolymer G (Glycine)



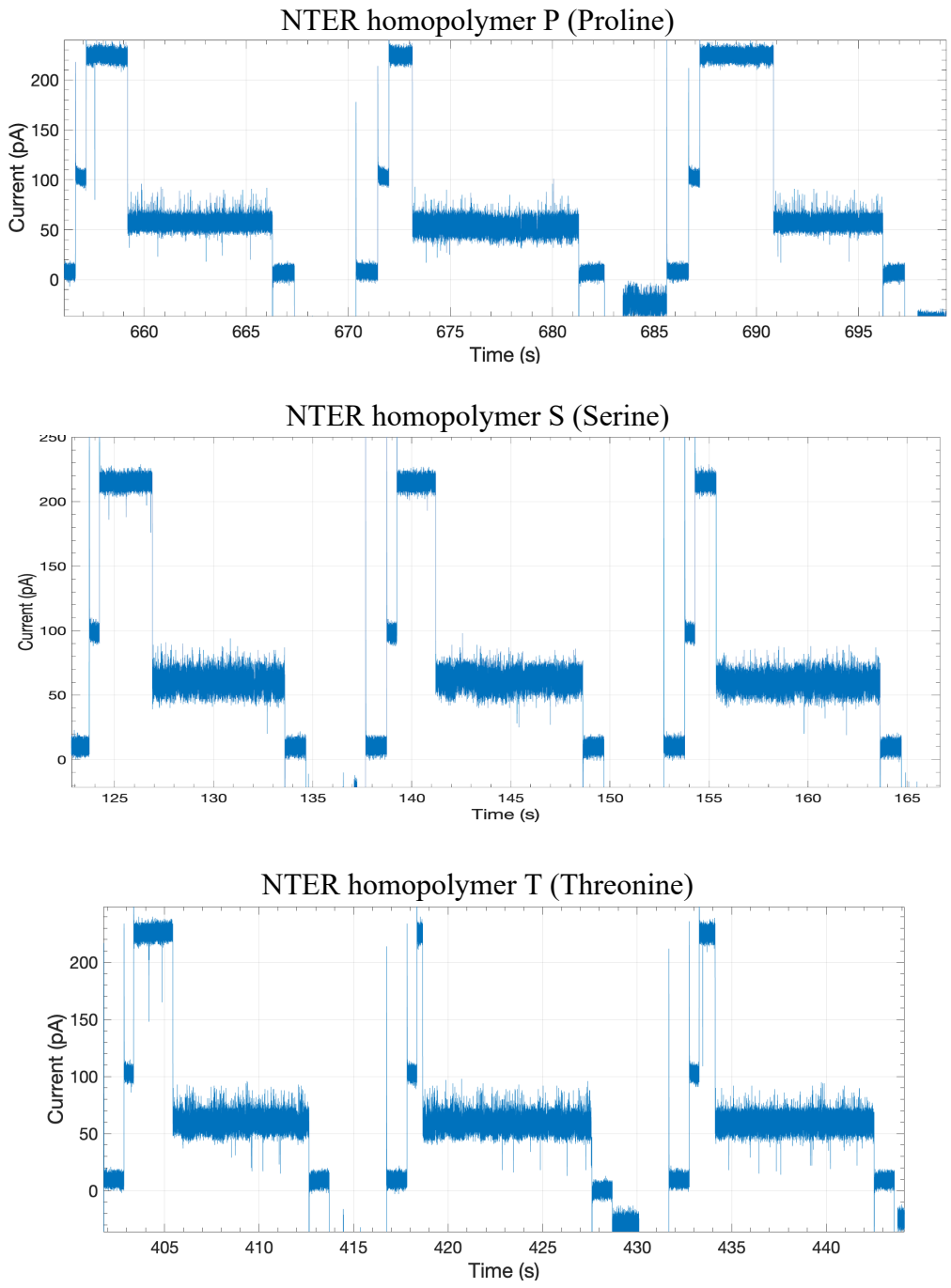
NTER homopolymer H (Histidine)



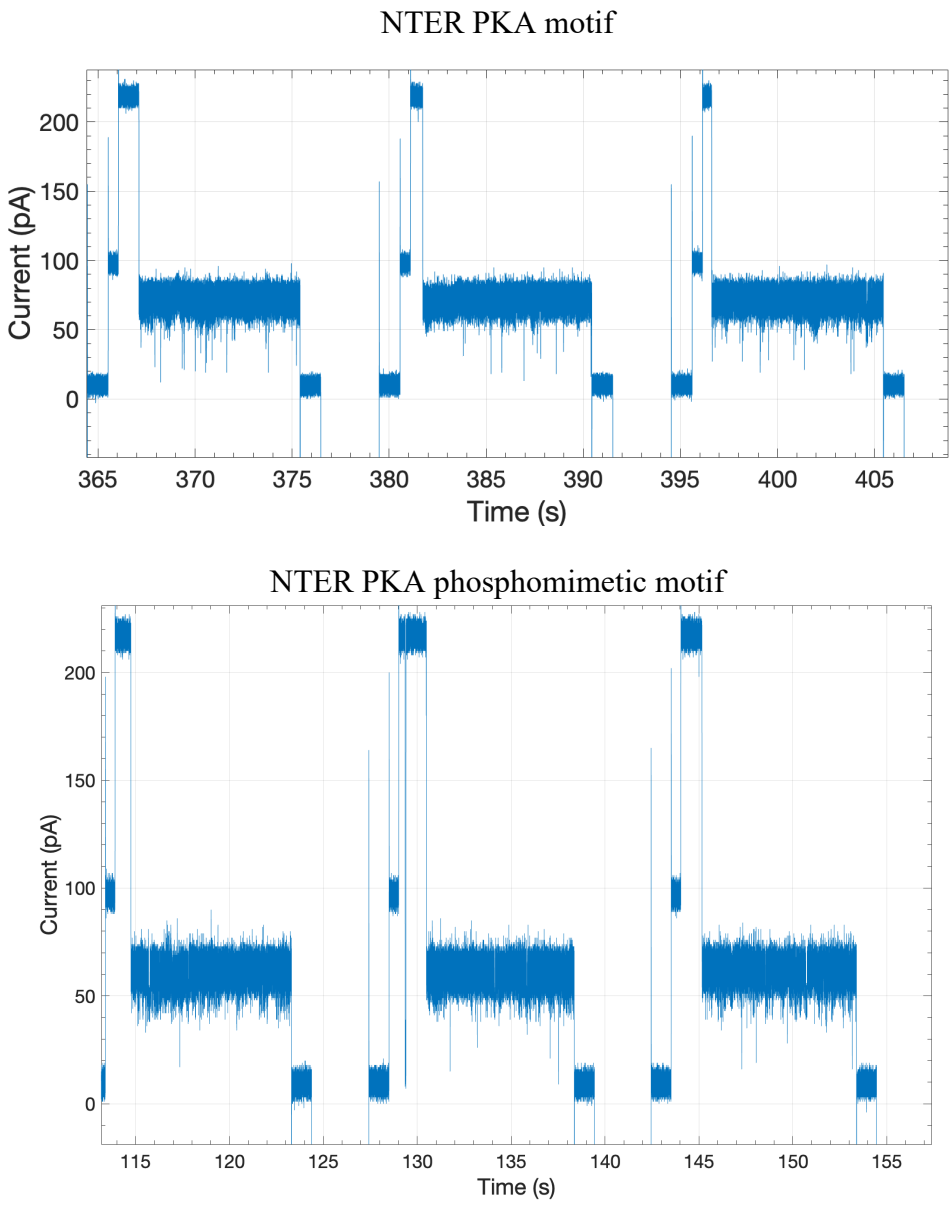
NTER homopolymer M (Methionine)

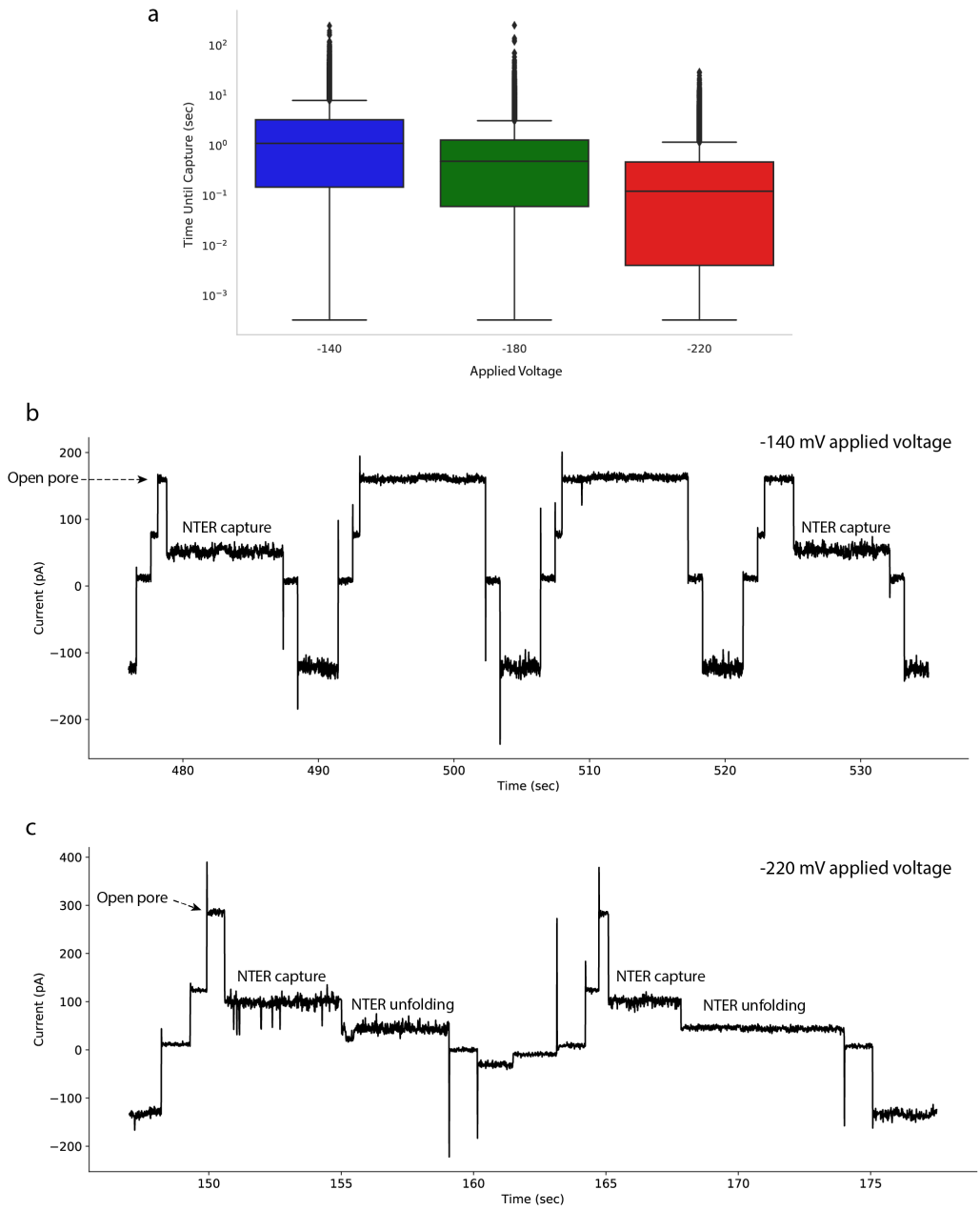


Supplementary Figure 3 (cont.)

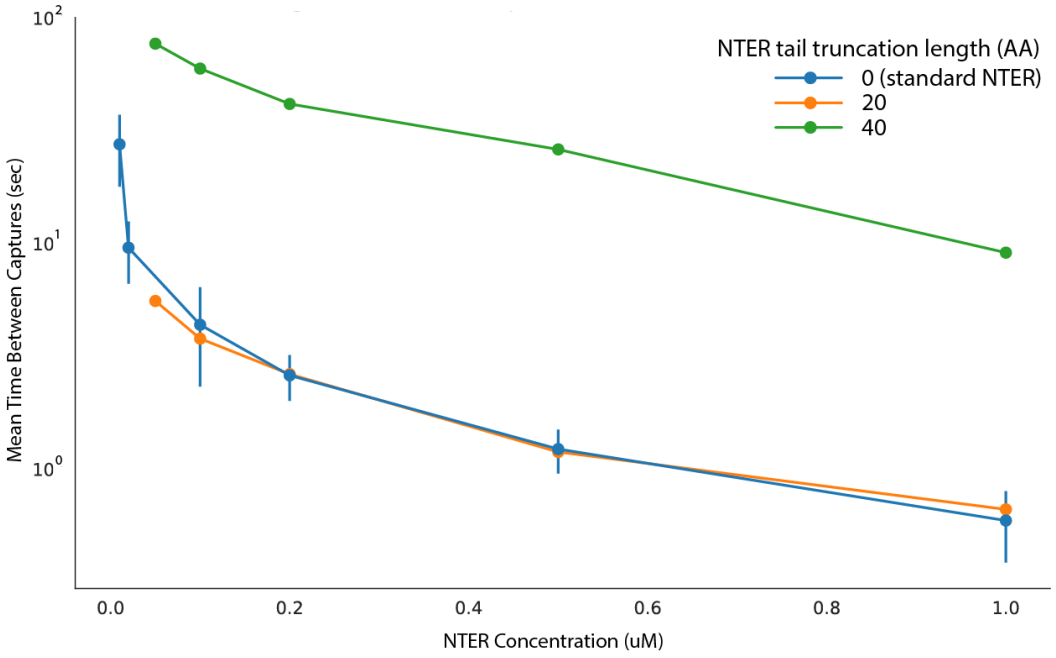


Supplementary Figure 3 (cont.)

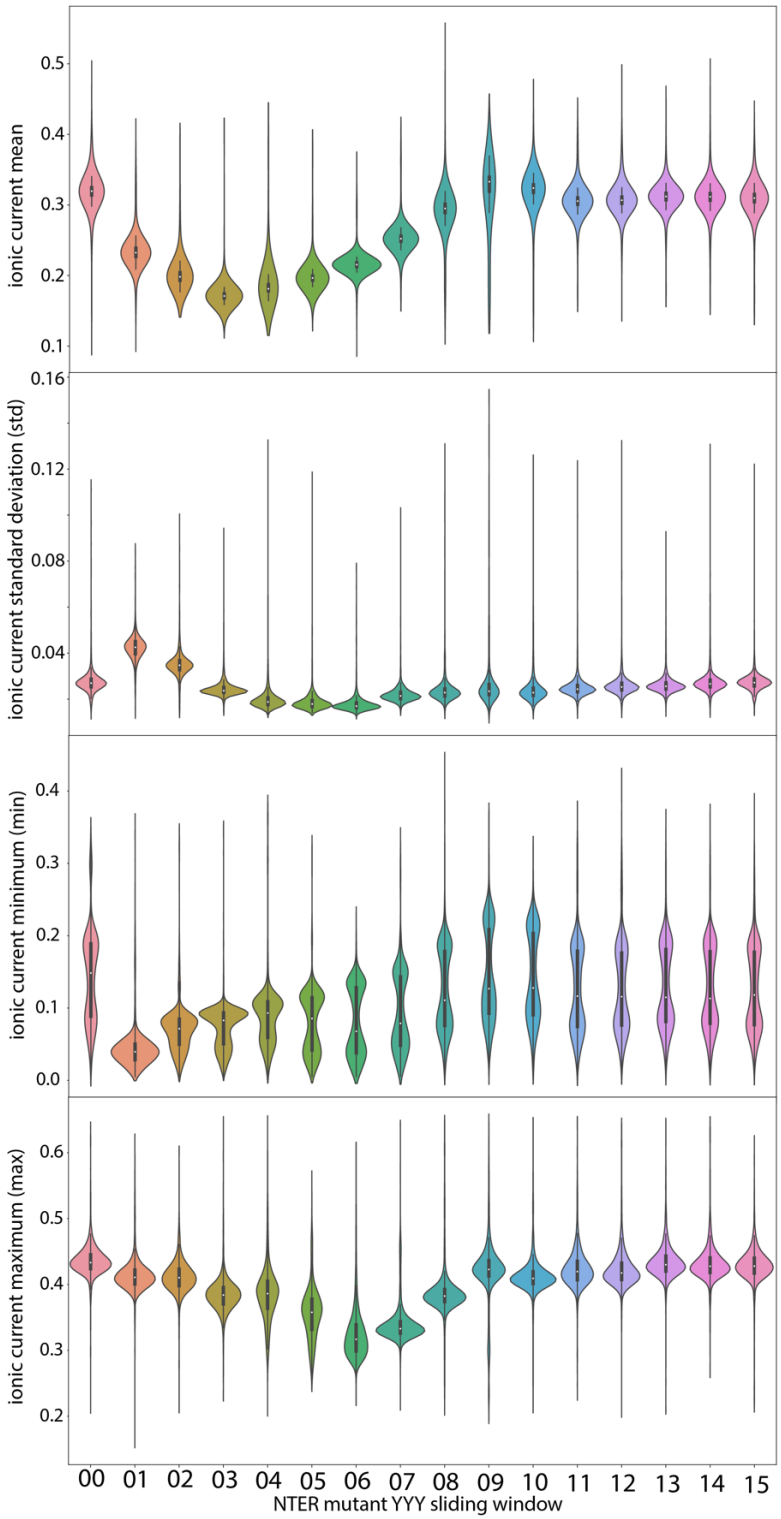




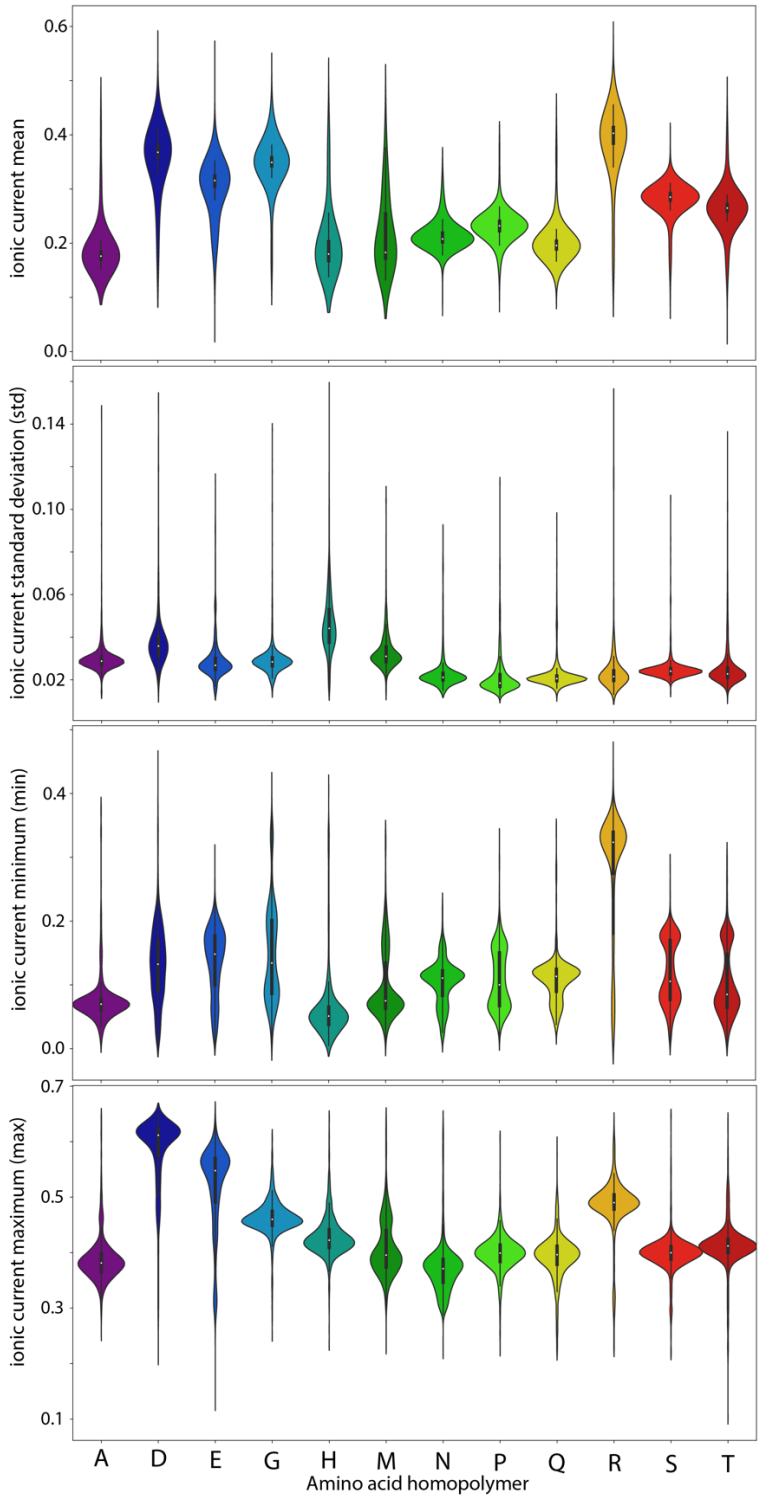
Supplementary Figure 4. The effect of the applied voltage on NTER capture rates. **a**, The "time between captures" data are presented as a box-and-whisker plot (center line: median, box: 1st and 3rd quartiles, whiskers: min and max, dot: outliers). Results represent the average times collected from n=3 independent nanopore experiments conducted with NTERs Y00-08 at 0.5uM. Example NTER captures at **b**, -140 mV, and **c**, -220 mV applied voltage. Increasing capture rates with higher applied voltages could be advantageous for this reporter system in the future, however, we do note that blockade events collected at -220 had a higher tendency to transition into a secondary blockage state following the initial capture state. We attribute this secondary state to unfolding of the Smt3 domain within the pore under the higher applied voltage force. Unfolding is not ideal, as it obscures analysis of the NTER barcode region in the pore. The signal analysis filtering step (see Methods) typically removes captures in which unfolding has occurred.



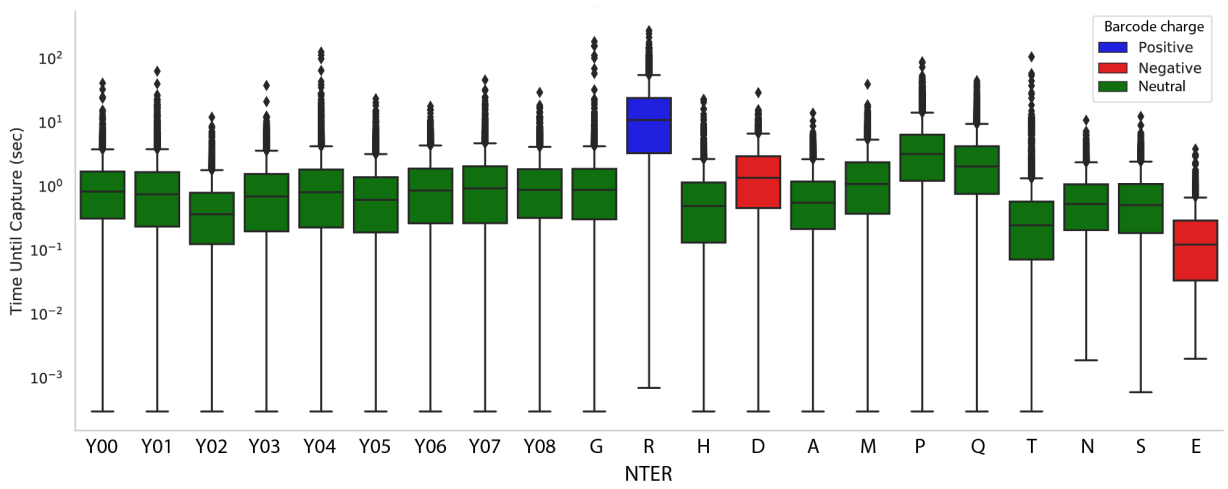
Supplementary Figure 5. The effect of NTER tail length on nanopore capture rates across different concentrations. Standard NanoporeTER tail sequence, 0 (blue). Tail sequences that have been truncated by either 20 (orange) or 40 (green) amino acids. Error bars represent standard deviation of the mean from n=3 independent nanopore experiments (standard NTER).



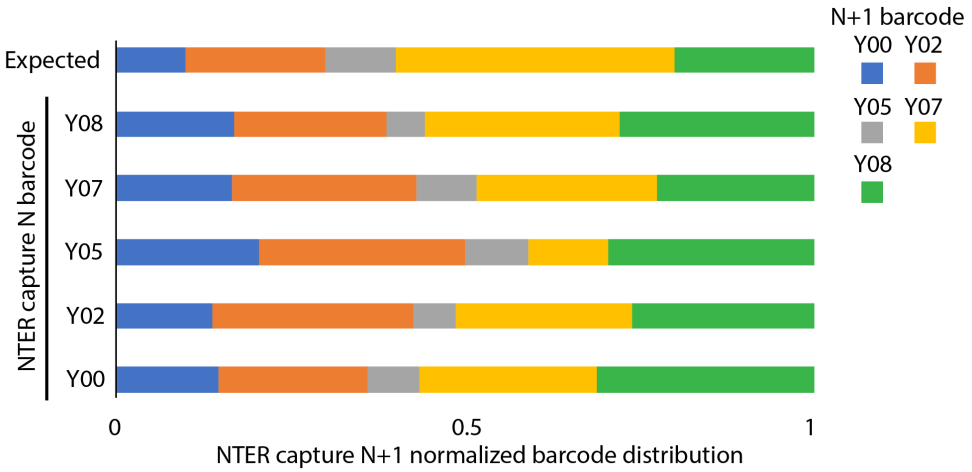
Supplementary Figure 6. Violin plots showing the ionic current level signal characteristics (mean, std, min, and max. All normalized to the open pore level) of the nanopore capture state for NTERs 00-15. Center dot: median, black box: 1st and 3rd quartiles, whiskers: 1st and 3rd quartiles -/+ 1.5 interquartile range. Each NTER distribution is composed of ~4000 single-molecule measurements.



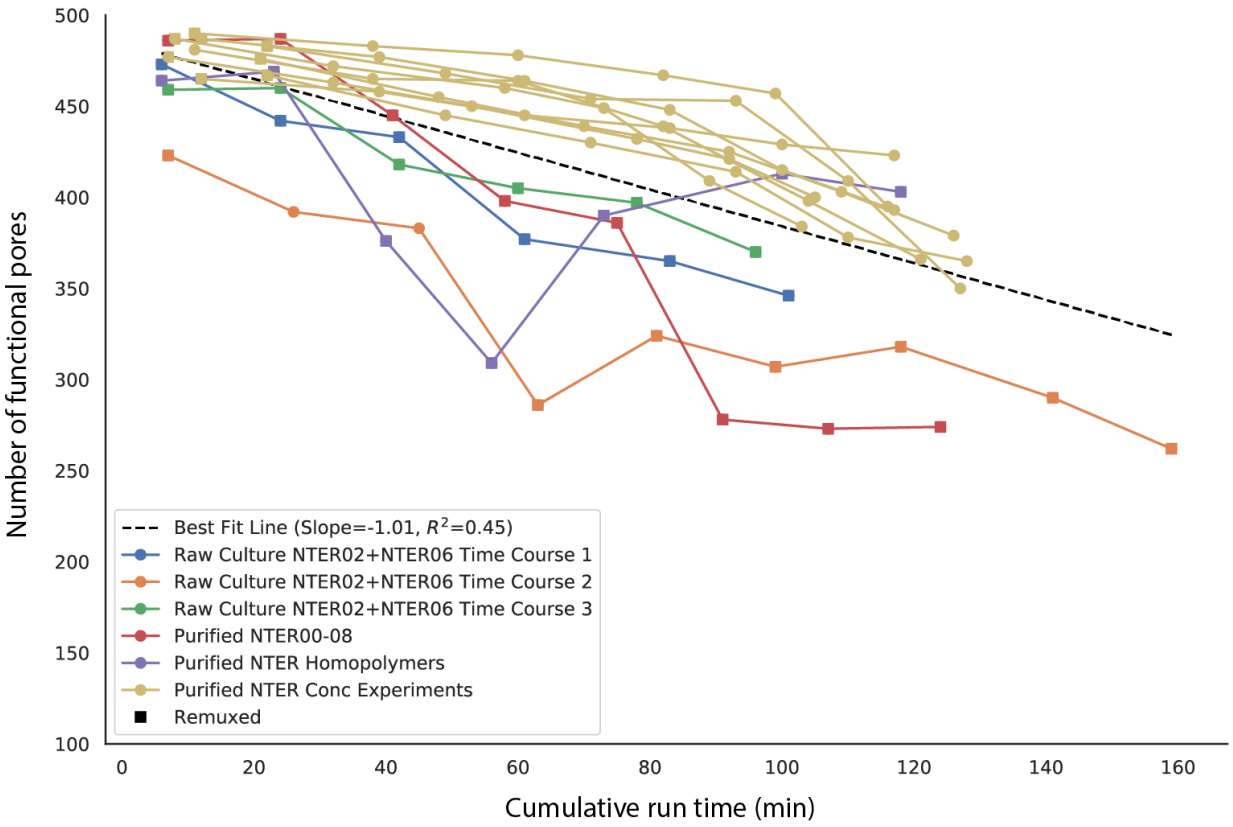
Supplementary Figure 7. Violin plots showing the ionic current level signal characteristics (mean, std, min, and max. All normalized to the open pore level) of the nanopore capture state for the amino acid homopolymer mutants. Center dot: median, black box: 1st and 3rd quartiles, whiskers: 1st and 3rd quartiles -/+ 1.5 interquartile range.Each NTER distribution is composed of ~1500 single-molecule measurements.



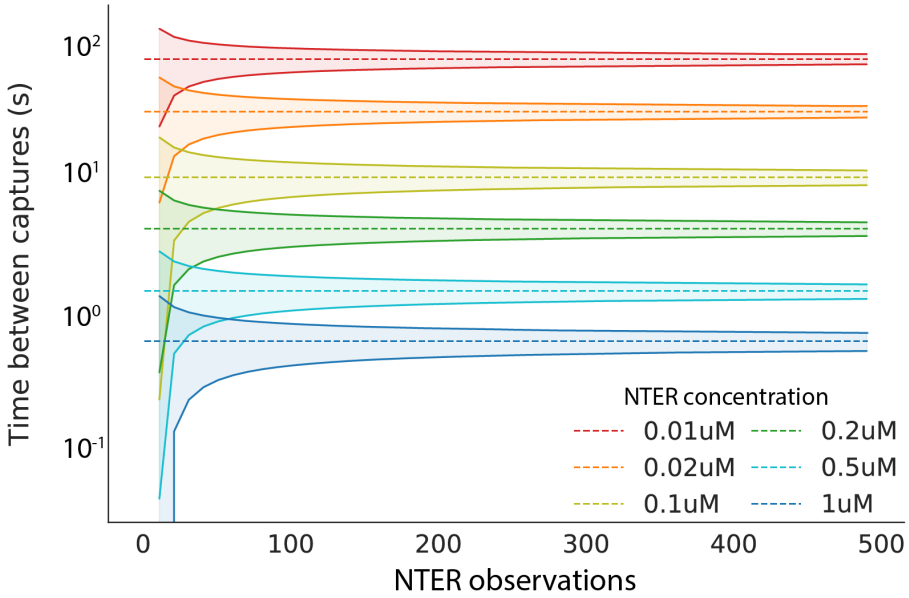
Supplementary Figure 8. The effect of NTER barcode sequence on nanopore capture rates at 0.5 uM concentration. The "time between captures" data are presented as a box-and-whisker plot (center line: median, box: 1st and 3rd quartiles, whiskers: min and max, dots: outliers). Barcodes with homopolymer charged amino acids are highlighted. Each NTER distribution is composed of ~4000 or ~1500 single-molecule measurements, for NTERs Y0-Y08 and homopolymer barcodes, respectively.



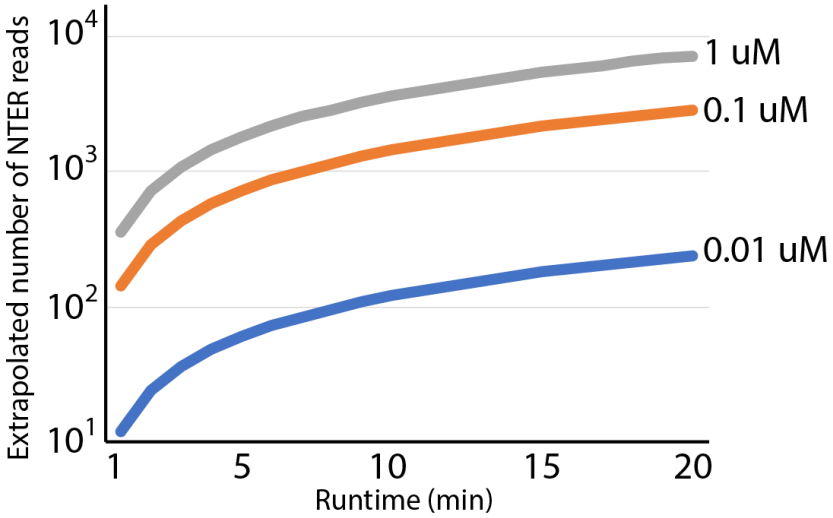
Supplementary Figure 9. Immediate NTER re-capture analysis. The bar plot shows the fraction of each barcode capture in a pore, in comparison to the expected fractions given 5 different NTER barcodes were mixed at varying concentrations. Y00: 0.05uM, Y02: 0.1uM, Y05: 0.05uM, Y07: 0.2uM, and Y08: 0.1uM. For each NTER barcode capture (N), we then determined distribution of the following captured NTER's barcode (N+1).



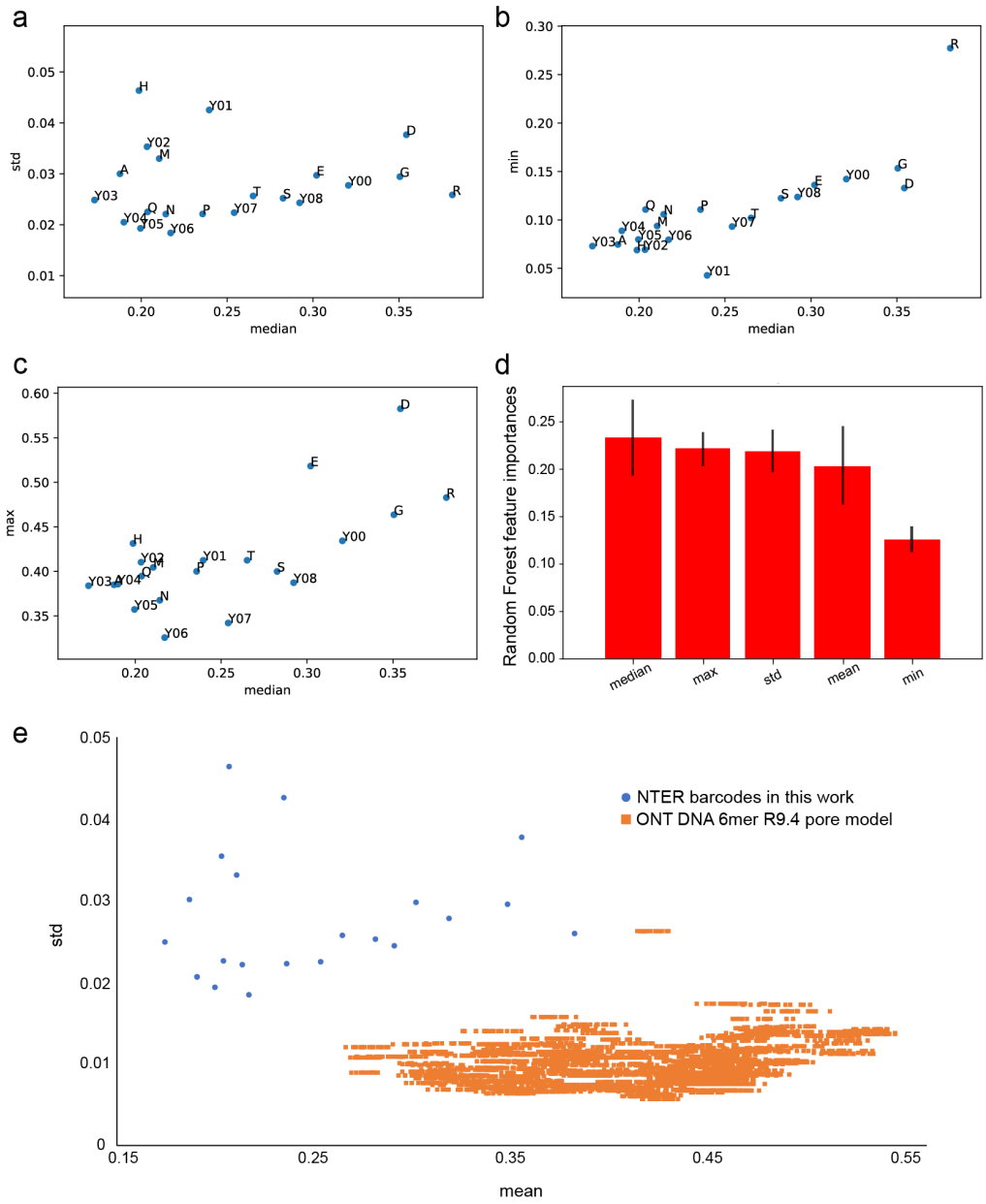
Supplementary Figure 10. MinION flow cell lifetime. The number of nanopores determined to be functional at the start of each experiment vs. cumulative flow cell runtime. Each line is a unique flow cell. Each point represents the start of a new experiment. Square points denote if the flow cell was re-muxed prior to the start of the experiment. Colors denote different sample types and experimental conditions.



Supplementary Figure 11. 95% confidence intervals for the mean time between peptide captures at multiple concentrations, varying the number of observed captures. NTERs Y00-Y08 were run at concentrations of 0.01 µM, 0.02 µM, 0.1 µM, 0.2 µM, 0.5 µM, 1 µM, then classified. We then calculated the time between peptide captures for each run, and pooled these values to combine all runs for the same concentration. The dashed line represents the mean time between captures for the runs pooled by concentration, and the shaded region represents the 95% confidence interval based on the observed mean and standard deviation, for varying numbers of observed NTERs. Note: this plot counts sequential NTER captures, meaning you have already observed at least one capture prior to these calculations.



Supplementary Figure 12. Relationship between NTER concentration and MinION run time at various NTER concentrations. These values are extrapolated from concentration experiments conducted on purified NTERY00, which had mean read rates of 361, 144, and 12 reads/min at NTER concentrations of 1 uM, 0.1 uM, and 0.01 uM (respectively), and an average of ~436 functional pores over the course of the experiments.



Supplementary Figure 13. NTER barcode signal space. **a**, Scatter plot showing the mean of the ionic current level signal characteristics of NTER barcodes Y00-08 and homopolymer mutants median vs std, **b**, median vs min, and **c**, median vs max. Each NTER point is the mean of >1000 single-molecule measurements. **d**, NTER Y00-08 and homopolymer mutant barcode classification feature importance determined using a Random Forest model trained on the five extracted signal features (median, mean, std, min, and max). Error bars represent the standard deviation of the feature importances for all trees in the forest, with n=300 trees. **e**, NTER Y00-08 and homopolymer mutant barcodes mean vs std (blue) compared to Oxford Nanopore Technology's DNA 6mer R9.4 pore model values (level_mean vs level_std) for every possible DNA 6mer (orange). (https://github.com/nanoporetech/kmer_models). The ONT pore model raw current values were converted into fractional current by normalizing them by a typical R9.4 open channel ionic current level of 220 pA.

Supplementary Notes

NTER protein design. We based our initial NTER design on the synthetic protein construct 'S1', which we previously developed for unfoldase-mediated nanopore analysis[12,13]. S1 contains a small, folded domain (Smt3) along with a flexible, negatively-charged 67 amino acid C-terminal 'tail' composed of glycine, serine, and acidic amino acid residues, in addition to an 11 amino acid ssrA tag[25]. The tail's lack of structure and net negative charge promotes capture of the protein in a nanopore sensor under an applied voltage. The ssrA tag allows for ClpX-mediated unfolding and translocation of the Smt3 domain, which otherwise inhibits translocation of S1 through the nanopore. For use as a reporter protein in *E. coli*, we modified protein S1 in two ways (**Figure 1a** and **Supp. Figure 1**). First, we replaced the ssrA tag with additional glycine/serine/acidic residues to preserve its nanopore threading activity but prevent targeting of the protein for degradation by ClpXP in-vivo. Second, we added an N-terminal OsmY domain[14]. In *E. coli*, OsmY-tagged proteins are secreted into the extracellular medium[15]. We reasoned secretion would facilitate NTER nanopore analysis by avoiding the need to lyse cells, thereby simultaneously reducing both experimental labor and signal noise that could be generated by non-specific interaction of intracellular molecular species (e.g. DNA, RNA, and other proteins) with the nanopores during analysis. Experiments in BL21 (DE3) *E. coli* showed that expression of this modified version of S1, which we term here 'NTERY00', resulted in secretion of the protein into the medium, as detected by SDS-PAGE analysis (**Supp. Figure 2**).

NTER sequence-to-signal relationship. We constructed NTER variants in which positions 3-11 within the polyGSD region were mutated to all the 20 possible standard amino acid homopolymers. **Figure 1k** and **Supp. Figure 7** show the signal features of the ionic current levels for 12 out of the 20 NTER homopolymer mutants (the homopolymers C, F, I, K, L, V, W, and Y, most of which have significant hydrophobic character, did not express sufficient soluble protein). To see how the different amino acid physical properties contribute to the NTER ionic current, we investigated whether certain properties correlate with different signal features. While no strong correlations were found across all the 12 amino acid types, we did find that the median current level moderately correlated with both amino acid volume and helical propensity within the uncharged amino acid types ($R = \sim 0.75$ for each, **Figure 1l,m**). We also considered if the barcode sequence could have an effect on the NTER capture efficiency into the nanopore. While we found that most of the sequences we tested had similar capture rates, the largest difference was observed for the positively charged Arginine homopolymer barcode, which captured ~10 fold slower than the other barcodes (**Supp. Figure 8**). To probe the potential of this method to resolve between amino acid barcodes with subtler sequence differences (for example, point mutations or post-translational modifications), we cloned and tested two additional NTER barcodes based on the eukaryotic protein kinase A (PKA) phosphorylation motif[17]. The first PKA-based barcode contained a canonical PKA motif (RRGSY), while the second had a single amino acid difference (RRGEY) that mimics the PKA motif's phosphorylated serine state in structure and charge (commonly referred to as a 'phosphomimetic', **Figure 1n**). Following purification and MinION analysis of these two NTERs, we found that the phosphomimetic barcode could be distinguished from the canonical PKA motif barcode, as the two barcodes typically had substantially different nanopore ionic current state medians (**Figure 1n**). These results suggest the potential of using NanoporeTERs to report on the activity of enzymes that regulate specific post-translational modifications, such as phosphorylation and methylation.

NTER experimental throughput. We assessed the experimental throughput of NTER protein measurements on a MinION, as they go beyond the platform's normal mode of operation. We found that experiments performed on new R9.4.1 flow cells typically started with >450 functional pores (**Supp. Figure 10**). Pores became non-functional (e.g. permanently clogged or ruptured bilayer) at a rate of ~1 pore/minute over the first two hours of flow cell lifetime, which was typically the longest amount of time we used a single flow cell. The usual MinION runtime for each of our experimental analyses was ~10 minutes per sample, along with an additional 5-minute wash step between samples (see **Methods**). Thus, close to 10 experiments could be performed per flow cell while maintaining a high number of functional pores (>300). We note, however, that the required runtime for each experiment is dependent on NTER concentrations, as lower concentrations require longer runtimes in order to collect enough observations to determine the average capture rate confidently. For example, to determine the approximate runtimes required to confidently arrive at the true mean NTER time between captures at varying concentrations, we computed the 95% Confidence Interval (CI) with respect to the number of observed captures (**Supp. Figure 11**). Overall, we found that a few hundred NTER reads was sufficient for CI convergence over the two orders of magnitude concentration range we tested (1 uM to 0.01 uM). For high concentrations, this number of reads can be collected in ~1 minute. While, at lower concentrations, it would take on the order of tens of minutes based on our extrapolations (**Supp. Figure 12**). Precisely how these numbers relate to absolute expression levels remains to be elucidated as more work will be needed to determine how NTER measurements correlate to absolutely quantitative expression measurements such as those that can be obtained with RNA and Ribo-Seq methods[26-28]. Finally, although our current NanoporeTER classification pipeline is currently done post-runtime, real-time results will be possible with the addition of software that runs concurrently (or within) the MinION operating scripts[29].

NTER barcode space. While we have characterized here a set ~20 orthogonal NanoporeTERs, examination of the sparsity of the current NTER signal space suggests that finding additional orthogonal barcodes is achievable (**Supp. Figure 13**). Although it is difficult to confidently estimate an upper bound to this signal-orthogonal barcode space at this juncture, in comparing the current NTER signal space to the DNA signal space of an R9.4 MinION flow cell (which has a sequence space of 4^6 unique kmers), we see that our current set of NTERs occupy a similar or even larger total area of the ionic current level vs std feature space (**Supp. Figure 13**). The potential to push this space further is reasonable considering that the homopolymer mutants only spanned 9 (residues 3-11) out of the 17 total positions that contribute to the barcode's ionic current signature. We are also able to consider more signal features than the DNA kmer model (owing to the long dwell times we are able to achieve by stalling the NTER barcode in the pore), such as the signal minimum and maximum. These additional signal features are useful for classification, as shown by the Random Forest classifier's feature importances (**Supp. Figure 13**). Signal features beyond this limited set can be manually extracted or learned by deep learning models that operate on the raw signal, such as our CNN approach, which is perhaps more scalable as the number of barcodes increases. This is somewhat analogous to the advances in nanopore sequencing of DNA achieved with deep learning-based models[30]. One challenge will be how to efficiently search the theoretical NTER barcode space given the vast number of possible sequences (~20^{17}). Ultimately, barcode space could be expanded through several different strategies, including: 1) high-throughput methods to empirically characterize more barcode sequences for classifier training, 2) engineering

NanoporeTERs to contain multiple barcode regions that can be consecutively read out with the aid of processive motor proteins[12,13] or voltage-mediated translocation[31], which would allow the number of orthogonal NTERs to scale exponentially with the number of individually characterized barcodes, and 3) semi-supervised machine learning models trained to accurately predict the sequence of empirically uncharacterized NTER barcodes given only their nanopore signal[32].

NTER secretion and measurement considerations. NanoporeTERs should not be limited to use in *E. coli* and *HEK293* cells, as many different N-terminal secretion domains have been characterized in a range of diverse organisms[33-35]. Relatedly, consideration will have to be given to the extra energy burden that NTER secretion puts on the cell, as well as the potential for secretion efficiency to be perturbed under different conditions. The higher energetic demands of secretion could be a potential drawback of the secretion approach. However, we note that secretion, while non-invasive and convenient, is not a necessary feature for future NanoporeTER designs. NanoporeTERs can be designed to lack the secretion domain. Minimal affinity-based purification (e.g. pulldown with magnetic beads or spin column) could be used to purify non-secreted NTERs directly from cell lysates. One other consideration to note is that cells could undergo a shift in secretion capacity under some experimental conditions that could result in a bias in the NTER measurements. One potential solution for correcting for potential shifts in secretion capacity would be use of a "reference" NTER under the control of a constitutive promoter that is expressed at a constant level. Expression levels of the experimentally-dependent NTERs could then be measured relative to the reference NTER. This would make it possible to identify, and potentially correct for, shifts in secretion capacity. Correlating the reference NTER nanopore-measured expression level to an absolute measurement, such as RNA-Seq or Ribo-Seq, could also be a way to obtain absolute measurements of intracellular NTER production rates. Also, as we observed that some barcode sequences have the potential to affect expression levels and/or nanopore capture rate, new barcodes should be similarly characterized for suitability.

Supplementary References

25. Baker, T. A. & Sauer, R. T. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim. Biophys. Acta - Mol. Cell Res.* **1823**, 15–28 (2012). doi: 10.1016/j.bbamcr.2011.06.007.
26. Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* (2014). doi:10.1016/j.cell.2014.02.033
27. Owens, N. D. L. *et al*. Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell Rep*. (2016). doi:10.1016/j.celrep.2015.12.050
28. Gorochowski, T. E. *et al*. Absolute quantification of translational regulation and burden using combined sequencing approaches. *Mol. Syst. Biol*. (2019). doi:10.15252/msb.20188719
29. Loose, M., Malla, S. & Stout, M. Real-time selective sequencing using nanopore technology. *Nat. Methods* (2016). doi:10.1038/nmeth.3930
30. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol*. (2019). doi:10.1186/s13059-019-1727-y
31. Rodriguez-Larrea, D. & Bayley, H. Multistep protein unfolding during nanopore

translocation. *Nat. Nanotechnol.* (2013). doi:10.1038/nnano.2013.22
32. Sutskever, I., Vinyals, O. & Le, Q. V. Sequence to Sequence Learning with Neural Networks. *Proc. NIPS*. 1–9 (2014).
33. Olczak, M. & Olczak, T. Comparison of different signal peptides for protein secretion in nonlytic insect cell system. *Anal. Biochem.* (2006). doi:10.1016/j.ab.2006.09.003
34. Bitter, G. A., Chen, K. K., Banks, A. R. & Lai, P. H. Secretion of foreign proteins from Saccharomyces cerevisiae directed by alpha-factor gene fusions. *Proc. Natl. Acad. Sci.* (2006). doi:10.1073/pnas.81.17.5330
35. Attallah, C., Etcheverrigaray, M., Kratje, R. & Oggero, M. A highly efficient modified human serum albumin signal peptide to secrete proteins in cells derived from different mammalian species. *Protein Expr. Purif.* (2017). doi:10.1016/j.pep.2017.01.003