# Label-free amino acid identification based on tRNA charging in a cavity and AMP blockade in a nanopore

*G. Sampath*
*P. B. 7849, J. P. Nagar P. O., Bengaluru-560078*
*Email: sampath_2068@yahoo.com*

**Abstract**. Label-free identification of amino acids (AAs) based on superspecific transfer RNAs (tRNAs) and amino-acyl tRNA synthetase (AARS) is explored. Molecules of a specific AA, a tRNA and its cognate AARS, and adenosine triphosphate (ATP) from a reservoir are confined to a micro-sized cavity by hydraulic pressure to counter the effects of diffusion. In this confined space a cognate tRNA in the cavity gets charged with AA and adenosine monophosphate (AMP) is released. The products are transferred to an electrolytic cell with a nanopore where AA, AMP, and ATP cause current blockades of different sizes. If an AMP-sized blockade occurs it means that a tRNA molecule has been charged with AA, this identifies AA; otherwise the procedure is repeated with a different tRNA and AARS until AA is identified. Unlike in most other nanopore-based methods, there is no need for precise measurement of current blockade levels. The efficacy of the procedure is assessed by simulating the movement of particles in a reservoir-cavity structure and analyzing the blockades that occur in the nanopore. In the former the probability of reactants escaping the cavity is found to be near 0. In the latter the probability of an error in identifying an AMP blockade is found to be less than 10% in the worst case (which occurs with the largest volume AA, namely Tryptophan). Detailed information and data are provided in a Supplementary File.

**Keywords:** amino acid; AMP; nanopore; label-free; AARS; current blockade; protein sequencing

## 1. Introduction

An established approach to sequencing of a peptide or protein is based on repeatedly cleaving the terminal residue at the N-end by Edman degradation (ED) [1] and identifying the cleaved residue with spectroscopic methods. Alternatively proteins may be broken into peptides for sequencing with mass spectrometry [2]. Both of these are bulk methods. Often test samples may contain only trace amounts so single molecule methods are of interest. With nanopores a single polymer molecule (DNA, RNA, protein) is electrophoretically transported through a pore in an electrolytic cell (e-cell) and the resulting current blockades probed to identify the sequence of monomers in the polymer [3,4]. A comprehensive review that charts developments in protein sequencing along four different directions is given in [5].

Recent work has focused on recognizing single amino acids that are free or are residues within a peptide. In [6] a nanopore is used to distinguish homopolymers of different lengths based on a single amino acid. In [7] it is reported that 13 of the 20 standard amino acids can be identified based solely on the volume excluded by an amino acid as it translocates through a nanopore. In [8,9] standard mass spectrometry methods are used to sequence monomers that are ionized after they are photo-cleaved from a polymer as it emerges from a nanopore.

Here a method of identifying free amino acid molecules is proposed and simulated in part. It combines nanopore-based measurements with chemical reactions that resemble those in the *in vivo* mRNA-to-protein translation process [10]. At its center is the superspecificity property of tRNAs: a tRNA molecule is almost always charged with its cognate amino acid [11]. The *in vivo* error rate without error correction is about 1 in 250, with error correction about 1 in 10000 [12]. As a result the extreme precision required with nanopores to measure blockade current levels due to 20 different monomer types [4], many differing only marginally from others, is no longer required. In the present case it is sufficient to distinguish the blockade level of *any amino acid* from the significantly larger one due to AMP (adenosine monophosphate), and of AMP from that of ATP. With a volume ratio of 1:1.5 between Tryptophan, the largest volume amino acid, and AMP, and a volume ratio of 1:1.25 between AMP and ATP, AMP blockade levels can be reliably distinguished from the others with currently available detector circuits.

This approach requires the use of a small cavity where tRNA charging can occur without the adverse effects of diffusion. With a source chamber and a nanopore on one side of the cavity and a reservoir on the other, single polymer molecules can be trapped inside the confined space for analysis or for desired chemical reactions. Such a structure was introduced in [13,14] and modified in [15] for the analysis of DNA. In the latter the trapped DNA is fragmented in the cavity (or 'cage') by a restriction endonuclease that enters the cage from the reservoir. The fragments are allowed to translocate through the pore back into the source chamber and the resulting current blockades used to study the fragments.

The method presented here uses a somewhat similar structure. It takes the form of a micro-sized container with a reservoir on one side from which the reactants enter the cavity/cage through a connecting passage based on a combination of diffusion and hydraulic pressure. After a suitable incubation time (to be determined experimentally)

the products are transferred to the *cis* chamber of an e-cell with a nanopore, where the resulting blockade levels are used in conjunction with the identity of the tRNA to identify the amino acid. The reservoir-cavity structure is simulated computationally to show that the reactants remain confined in the cavity so that charging of a tRNA with its cognate AA can occur before the products are transferred to the e-cell.

The efficacy of the approach is assessed by: 1) computationally simulating particle movement in the reservoir-cavity structure; and 2) estimating the error in distinguishing among blockades due to AA, AMP, and ATP in the e-cell. In 1 the probability of reactants escaping from the cavity is found to be near zero, in 2 the error rate is < 10%. In conjunction with superspecificity these results suggest that the proposed method can accurately identify amino acid identification without any kind of labeling.

## 2. Method

There are two stages: 1) Charging of tRNA; 2) Amino acid identification based on current blockades in a nanopore. The first step is modeled after aspects of the mRNA-to-protein translation process, the second is an *ad hoc* addition.

**Stage 1**: *Charging of tRNA with an amino acid (AA)*

In a biological cell an amino-acyl tRNA synthetase (AARS) enzyme activates ATP and a cognate AA to form an amino acid AMP complex, charges the cognate tRNA with AA at the latter's C-terminal end and releases the AMP, then releases the charged or aminoacyl tRNA [10]. tRNA superspecificity [11] means an error rate of about 1 in 250 without error correction and 1 in 10000 with error correction [12]. The following equations summarize the process:

$$\text{AA} + \text{ATP} \rightarrow \text{Aminoacyl-AMP} + \text{pyrophosphate or PPi } (P_2O_7^{4-}) \quad (1)$$

$$\text{Aminoacyl-AMP} + \text{tRNA} \rightarrow \text{Aminoacyl-tRNA} + \text{AMP} \quad (2)$$

$$P_2O_7^{4-} + H_2O \rightarrow 2\,HPO_4 + 2\,H^+ \quad (3)$$

This process can be used *in vitro* to charge a tRNA with a free AA. When the latter is exposed in succession to distinct tRNAs, their cognate AARSs, and ATP in solution, one of the tRNAs will get charged with AA. This may be done up to N ($20 \leq N \leq 61$) times with N different tRNAs. Here 20 is the number of standard AAs, and 61 the number of coding triplets in the genetic code. 'Wobble' in the third position [10] does not matter as only one tRNA type is used in each of the 1 to N cycles. With N = 20, the minimum number of cycles required is 1, the maximum 20, and the average 10.

The reactants need to be restricted to a small space because otherwise diffusion takes over. Such restricted spaces may take the form of sub-micro- to nano-volume containers or cavities. With a reservoir connected to a cavity, polymer molecules from the former can be trapped inside the latter. In [13,14] a pore-cavity-pore structure was used in the study of single molecules. This was modified in [15] for the analysis of long DNA strands. Other related structures are described in [16-18].

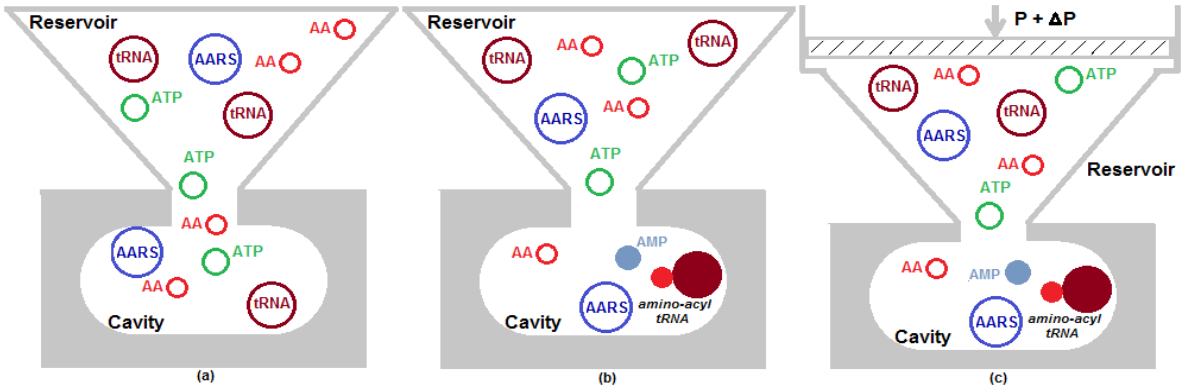


**Figure 1**. Schematic diagram (not to scale) showing charging of tRNA with cognate amino acid. tRNA. AARS, and ATP molecules (Equations 1 and 2) are added to buffer solution containing AA molecules in reservoir. Taper in reservoir increases probability of particle translocating toward cavity. Reactants enter cage by diffusion and hydraulic pressure. (a) tRNA not cognate to AA, not charged. (b) tRNA and AARS cognate to AA, AARS breaks up ATP molecule into AMP molecule and pyrophosphate (not shown), AMP provides energy for charging, amino-acyl tRNA formed in cavity. (c) Hydraulic pressure $P_{atm} + \Delta P$ (where $P_{atm}$ = atmospheric pressure and $\Delta P$ is the applied pressure) provides bias in z direction toward cavity bottom.

*Reservoir-cavity structure*

The approach described here is based on a modified form of the structure in [15]. A schematic diagram is shown in Figure 1. One or more free molecules of a specific AA, tRNA, AARS, and ATP enter a micro-sized cavity from a tapered reservoir through a sub-micro-sized passage (Figure 1a). The taper ensures that molecules in the reservoir translocate toward the cavity, where one or more cognate tRNAs can get charged with AAs (Figure 1b). The passage from the reservoir to the cavity must be large enough to allow AA, AMP, ATP, tRNA, and AARS to pass through. As a result tRNA charging can occur in a confined space without the dispersive effects of diffusion. Table 1 shows the sizes of the reactants (see Supplementary File for more information). The dimensions of the reservoir, passage, and cavity are on the same order as in [13-15]. No measurements are made so they do not have to be exact.

**Table 1**. Sizes of reservoir-cavity structure and reactants involved in tRNA charging. Amino acids are represented by smallest (Glycine) and largest (Tryptophan).

| Space or reactant | Dimensions of box (or ellipsoid axes) enclosing structure (or reactant) | | | Reactant | Axes of ellipsoid enclosing reactant | | |
|---|---|---|---|---|---|---|---|
| | Length (nm) | Breadth (nm) | Height (nm) | | Major (nm) | Minor (nm) | Minor (nm) |
| Reservoir in Figure 1 | 10000 | 10000 | 10000 | AMP [3] | 1.67 | 0.95 | 0.4 |
| Passage in Figure 1 | 200 | 200 | 400 | ATP [4] | 2.34 | 1.09 | 0.44 |
| Cavity in Figure 1 | 2000 | 2000 | 1500 | Phe tRNA [5] | 10.56 | 6.54 | 3.64 |
| Glycine (AA) [1] | 0.56 | 0.35 | 0.21 | Ala AARS [6] | 13.94 | 9.41 | 8.46 |
| Tryptophan (AA) [2] | 1.10 | 0.7 | 0.34 | | | | |

[1,2,3,4,5,6] Based on minimum volume ellipsoid covering centers of all atoms in molecule; see Supplementary File.

*Confinement of reactants to the cavity*

A directed force field is needed to ensure that the reactants enter the cavity and remain there without diffusing back to the reservoir. An electrophoretic field is problematic as AMP, ATP, tRNAs, and some AAs carry a negative charge, other AAs have a positive, negative, or zero charge, and AARSs have a net positive, negative, or zero charge; all these values depend on the pH of the solution [19]. The solution therefore is to use a charge-independent force field, such as hydraulic pressure, which has been used in the study of polymer translocation through nanopores [20].

This latter approach is used in the present work; see Figure 1c. Simulation studies, briefly discussed in Section 3 and detailed in Supplementary File 1, show that the probability of reactants escaping from the cavity is close to 0.

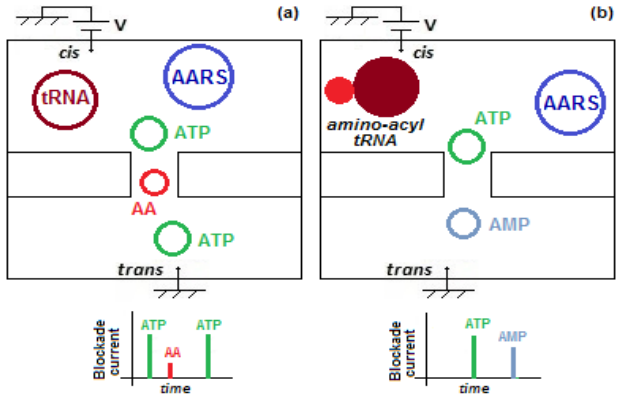


**Figure 2.** Schematic diagram of e-cell showing filtering effect of nanopore on tRNAs, AARSs, ATP/AMP, and AA. a) Uncharged tRNA remains in *cis* along with AARS, while free AA and two ATP molecules translocate through pore to *trans*; current blockades are due to free AA and ATP; b) Charged tRNA remains in *cis* along with AARS, while one ATP and one AMP molecule translocate through pore to *trans*; current blockades are due to ATP and AMP.

**Stage 2**. *Amino acid identification based on current blockades in a nanopore and identity of a charged tRNA*

In *in vivo* translation the AA attached to a tRNA is effectively identified for attachment to a growing peptide when the anticodon in the amino-acyl tRNA binds with a codon in an mRNA by Chargaff base pairing [10]. There is no

strict counterpart in the present case, so an *ad hoc* approach is used. Following charging of a tRNA (the wait time for which may be determined experimentally) the solution contents are transferred to the *cis* chamber of an e-cell with a nanopore where blockades due to analyte particles translocating through the pore when a voltage is applied between *cis* and *trans* can be observed (Figure 2). The size of the pore must be such that AA, AMP, and ATP can translocate via diffusion and electrophoresis to *trans* while tRNA, AARS, and amino-acyl-tRNA are blocked.

*Current blockade as a proxy for analyte volume*

Assume that the size of the current blockade $I_{AN}$ due to an analyte molecule AN is roughly proportional to the volume excluded when passing through the pore [7,21]. Then

$$I_{AN} \sim V_{AN} \tag{4}$$

can be used to identify the analyte molecule, which can be AA, AMP, or ATP. In what follows the volume of an analyte is used as a proxy for the current blockade that it causes. Table 2 shows the volumes (in $10^{-3}$ nm$^3$) of the 20 standard amino acids [22], and of AMP [23] and ATP, whose volume is approximated as 1.25 × AMP volume.

**Table 2.** Volumes of the standard 20 amino acids, ATP, and AMP. (Amino acid volumes are from [22], for AMP from [23]. ATP volume set to 1.25 × volume of AMP. All volumes in Å$^3$ or × $10^{-3}$ nm$^3$.

| **AA** | **G** | **A** | **S** | **C** | **D** | **T** | **N** | **P** | **V** | **E** | **Q** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Mean (μ)** | 59.9 | 87.8 | 91.7 | 105.4 | 115.4 | 118.3 | 120.1 | 123.3 | 138.8 | 140.9 | 145.1 |
| **AA / Other** | **H** | **M** | **I** | **L** | **K** | **R** | **F** | **Y** | **W** | **AMP** | **ATP** |
| **Mean (μ)** | 156.3 | 165.2 | 166.1 | 168 | 172.7 | 188.2 | 189.7 | 191.2 | 227.9 | 349.0 | 436.5 |

*Detecting the occurrence of tRNA charging and identifying an AA from an AMP blockade*

Of the three types, the 20 amino acids have the lowest volumes. The volume of AMP (349.0 × $10^{-3}$ nm$^3$) is at least 50% higher than that of the largest amino acid (W, Tryptophan, volume = 227.9 × $10^{-3}$ nm$^3$). The volume ratio between AMP and ATP is 1:1.25. Therefore the blockade due to AMP can be easily distinguished from the blockade due to any AA, and sufficiently reliably from that due to ATP, by setting two thresholds:

$$Th_{low} = (\mu_W + \mu_{AMP})/2 \tag{5a}$$

and

$$Th_{high} = (\mu_{AMP} + \mu_{ATP})/2 \tag{5b}$$

Additionally a lower threshold of $Th_{noise}$ is set to distinguish the blockade due to G from baseline noise. This leads to

$$Th_{low} < V_{AN} < Th_{high} \quad \rightarrow \quad AN = AMP \tag{6a}$$
$$Th_{noise} < V_{AN} < Th_{low} \quad \rightarrow \quad AN = AA \tag{6b}$$
$$V_{AN} > Th_{high} \quad \rightarrow \quad AN = ATP \tag{6c}$$

Crucially,

$$AN = AMP \quad \rightarrow \quad \text{a tRNA molecule has been charged} \tag{7}$$

If at least one AMP-sized blockade is observed

$$\text{Identity of AA} = \text{AA corresponding to charged tRNA} \tag{8}$$

If no AMP-sized blockade is observed the contents of *trans* are transferred back to the reservoir (Figure 1). A different tRNA and related AARS are added in up to N ($20 \leq N \leq 61$) cycles until AA is identified. As there is at least one tRNA cognate to an AA, some tRNA will get charged so the loop will always terminate.

Let there be $K_1$ molecules of free AA, $K_2$ ($\geq K_1$) of tRNA, $K_3$ ($\geq K_1$) of AARS, and $K_4$ of ATP ($\geq K_1$) in the micro-cavity. Let $K_2'$ of the tRNAs be charged. The solution transferred to the e-cell contains the following analytes:

$K_1$ - $K_2'$ free AAs; $K_2'$ amino-acyl tRNAs; $K_2$ - $K_2'$ uncharged tRNAs; $K_3$ AARSs; $K_4$ - $K_2'$ ATPs, $K_2'$ AMPs, and $K_2'$ pyrophosphates. With pore diameter of an appropriate size the tRNAs (charged or not) and AARSs will remain in *cis* because their tertiary structures are too big to enter the pore. Likewise the phosphate ions are too small to cause an appreciable blockade. Thus $K_1$ - $K_2'$ free AAs; $K_2'$ AMPs; $K_4$ - $K_2'$ ATPs will translocate through the pore and cause a total of $K_1$ - $K_2'$ + $K_4$ blockades. Figure 2 shows two cases with $K_1 = 1$, $K_2 = K_3 = 1$, and $K_4 = 2$, the first with $K_2' = 0$ and the second with $K_2' = 1$. These result in 3 and 2 blockades respectively.

**3. Simulating the reservoir-cavity structure**

After the reactants come together in the cavity they should remain there long enough for charging of a cognate tRNA to occur without diffusing away. To assess the efficacy of this the trajectory of particles in the reservoir-cavity structure was simulated with a range of parameters including diffusion coefficient, hydraulic pressure level, origin of particle, and number of diffusion steps. See the Supplementary File for a description of the geometry (Figure S-1), protocol, and computations involved.

The simulation procedure is similar to the procedures used in [24,25] and [26]. In [24,25] the terminal base of a DNA molecule is cleaved by an exonuclease enzyme, passes through a nanopore, and is identified based on the current blockade level it causes. In [26] the terminal residue of a peptide passing through the first pore of an e-cell with two nanopores in tandem is cleaved by an exopeptidase on the exit side of the pore, enters and translocates through the second pore, and is identified by the level of the blockade caused in the latter.

The simulation can be viewed in the following terms:

1) Each simulation step takes place in 1 picosecond ($10^{-12}$ s).

2) A reactant (free AA, AARS, ATP, AMP, tRNA, amino-acyl tRNA) is a dimensionless particle that is reflected off the walls of the reservoir, the passage, and the cavity.

3) The particle is considered to be confined when it reaches the bottom of the cavity and stays close to it within a specified margin for a specified time. This would suggest that the escape probability would be virtually 0.

4) The hydraulic pressure field superimposed by the piston on diffusion is assumed to result in Poiseuille flow in all three segments: reservoir, passage, and cavity [20]. The resulting velocity $v_{hyd}$ is given by

$$v_{hyd} = (P_{atm} + \Delta P)\, R^2_{channel} / 8\, \eta\, H_{channel} \qquad (9)$$

where $P_{atm}$ = atmospheric pressure, $\Delta P$ = applied pressure, $R_{channel}$ = radius of channel (reservoir, passage, or cavity), $\eta$ = solution viscosity, and $H_{channel}$ = radius of channel.

5) The displacement of a particle is equal to the sum of the vector of diffusive translation and the displacement due to hydraulic pressure (which is always in the z direction).

6) In the absence of hydraulic pressure the motion of the particle is entirely due to diffusion.

The following measures are used to assess the efficacy of confinement in the cavity: 1) Time of entry of a particle into the cavity; 2) Time to reach the bottom of the cavity; and 3) Dwell time at bottom of cavity. The simulation is stopped when the dwell time exceeds $10^6$ steps.

**Table 3**.

| Origin | $\Delta P$ (× atmospheric pressure) | Diffusion coefficient ($10^{-10}$ $m^2/s$) | $N_1$ = Time of entry into cavity (× $10^{-6}$ s) | $N_2$ = Time to reach bottom of cavity (× $10^{-6}$ s) | $N_3$ = Dwell time at bottom of cavity (× $10^{-6}$ s) |
|---|---|---|---|---|---|
| (0,0,0) | 0.5 | 0.3 | 0.800841 | 0.913141 | > 1.0 |
| | | 3.0 | 0.782680 | 0.895770 | > 1.0 |
| | | 30.0 | 1.038547 | 1.149302 | > 1.0 |
| (0,0,-$H_R$) | 0.5 | 0.3 | 0.997194 | 1.109616 | > 1.0 |
| | | 3.0 | 0.992062 | 1.104224 | > 1.0 |
| | | 30.0 | 1.272460 | 1.383287 | > 1.0 |
| ($W_R$/2, $W_R$/2, -$H_R$) | 0.5 | 0.3 | 1.079808 | 1.192158 | > 1.0 |
| | | 3.0 | 1.200182 | 1.311785 | > 1.0 |
| | | 30.0 | 1.099228 | 1.210150 | > 1.0 |

Table 3 shows the results for a subset of the parameters used in the full simulation, corresponding to the case when the hydraulic pressure = $P_{atm}$ + 0.5 atmospheric pressure. (The full set of results can be found in the

Supplementary File.) The last column suggests that if the simulation were to continue the particle would remain indefinitely near the bottom of the cavity. This of course is the desired outcome. These results may be compared with [13,14] where the confinement time is a few seconds, while in [15] the time is reported to exceed 25 minutes.

Table 4 shows the results for when the only motive force is diffusive (no hydraulic pressure is applied). In this case the measures used are: 1) Total number of diffusive steps; 2) The final value of z; 3) The maximum value of z attained by the particle (which need not the the same as the final value of z). The results indicate that with pure diffusion the particle enters the cavity in only one case, even then it does not reach the cavity bottom. In all three cases the particle is still diffusing in the reservoir at the end of the run. This shows that the use of hydraulic pressure is key to the confinement of reactants to the cavity so that charging of a tRNA to a cognate AA can take place.

**Table 4**. Diffusion coefficient = $3.0 \times 10^{-10}$ $m^2/s$

| Origin | $N_1$ = Total number of steps taken | Value of z after $N_1$ steps | Maximum z during $N_1$ steps | Entered cavity? | Reached bottom of cavity? |
|---|---|---|---|---|---|
| (0,0,0) | $4 \times 10^8$ | $-8.815216 \times 10^6$ | $0.0059 \times 10^6$ | Yes | No |
| $(0,0,-H_R)$ | $4 \times 10^8$ | $-9.750791 \times 10^6$ | $-9.6616 \times 10^6$ | No | No |
| $(W_R/2, W_R/2, -H_R)$ | $4 \times 10^8$ | $-9.901440 \times 10^6$ | $-9.6722 \times 10^6$ | No | No |

## 4. Error analysis

In analyzing the identification procedure in Section 3 for potential errors the following assumptions are made:

1) Superspecificity applies *in vitro* so that charging of a tRNA in the presence of a cognate AA always occurs and there are no charging errors. Correction of errors by AARS, as occurs *in vivo*, is not considered here.

2) If a tRNA and its cognate AA are not present then the AARS cognate to the tRNA does not activate ATP, which means that there is no AMP release (Equations 1 and 2).

The key step in the identification procedure in Section 3 is determining if the tRNA is charged or not. This determination depends only on the current blockade levels observed in the e-cell. With perfect identification the following two conditions hold:

a) The blockade level of AMP can always be distinguished from those due to any and all of the ATP molecules as well as due to any AA; and

b) When the tRNA is charged the released AMP causes a blockade of characteristic size.

In practice identification may not be perfect. Errors can be divided into *false positives* and *false negatives*. A false positive occurs when the tRNA is not charged but the output reads 'charged', while a false negative occurs when the tRNA is charged but the output reads 'not charged'. The probabilities of these two occurrences can be estimated by using $V_{AN}$ as a proxy for the measured current blockade $I_{AN}$ (Equation 4). Such analyses are usually simplified by assuming the relevant variables to be normally distributed with some mean $\mu$ and standard deviation $\sigma$. (This may be compared with the Voigt distribution in [24], a theoretical model based on a DNA sequencing method in which successive bases are cleaved and the blockades they cause when they translocate through a nanopore are measured [27]. The bases A, T, C, and G are distinguished on the basis of their volumes, which differ only marginally (see Table I in [23]).)

Let $F_X(x, \mu, \sigma)$ be the cumulative distribution function (cdf) of the gaussian variable X, the current blockade level for analyte X. With AN = AA, ATP, or AMP, an observed blockade is evidence of translocation of an amino acid AA if

$$\mu_{AA} - k_{AA}\,\sigma_{AA} \;<\; V_{AA} \;<\; \mu_{AA} + k_{AA}\,\sigma_{AA} \qquad (10)$$

and of ATP (AMP) if

$$\mu_{ATP(AMP)} - k_{ATP(AMP)}\,\sigma_{ATP(AMP)} \;<\; V_{ATP(AMP)} \qquad (11)$$

for some positive integer value of k times the spread of the distribution of analyte volume. The probabilities of false negatives and false positives are calculated next. The calculations are on the conservative side and use a spread of $5\sigma$ (k = 5) about the mean of the normal distributions involved.

1) *False negatives*

When a tRNA is charged, the tRNA is bound to the AA, and an ATP molecule breaks down into AMP and Ppi molecules. The AMP molecule translocates to *trans*. However the output signal reads 'not charged'. This misread

could result from the AMP causing an AA-sized blockade. Consider the example of Figure 2a with two ATP molecules. A false negative occurs if one of the ATP molecules causes an ATP-sized blockade and the AMP causes an AA-sized blockade. The probability of this is

$$P_{FN\text{-}AA} = F_{AMP}(Th_{low}, \mu_{AMP}, \sigma_{AMP}) - F_{AMP}(\mu_{AA} - 5\,\sigma_{AA}, \mu_{AMP}, \sigma_{AMP}) \quad (12)$$

Figure 3 shows the variation of $P_{FN}$ for each of the 20 amino acids for four values of σ (0.05μ, 0.1μ, 0.15μ, 0.2μ).

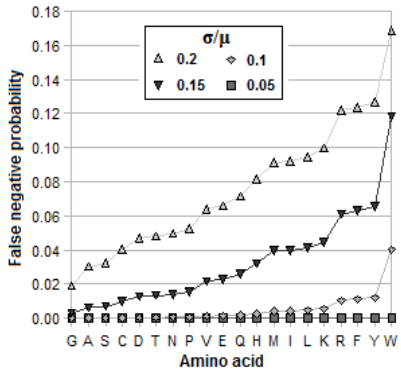


**Figure 3**. False negative probabilities for each of the standard 20 AAs. AA volumes are normally distributed with mean μ as given in Table 2. Probabilities are calculated for four values of standard deviation σ relative to μ.

2) *False positives*

In this case the tRNA is not charged, which means that there is a free AA and no reduction in the number of ATP molecules. However the signal says 'charged'. In this misread signal it would appear as if an AMP translocated to *trans* (along with other free ATP molecules in the solution), or possibly one of the ATP looks like an AMP and the AMP looks like an ATP. (The latter has a much smaller probability and need not be considered.) The probability of this is

$$P_{FP\text{-}AA} = F_{AA}(Th_{high}, \mu_{AA}, \sigma_{AA}) - F_{AA}(Th_{low}, \mu_{AA}, \sigma_{AA}) \quad (13)$$

Figure 4 shows the variation of $P_{FP}$ for each of the 20 amino acids for four values of σ (0.05μ, 0.1μ, 0.15μ, and 0.2μ).

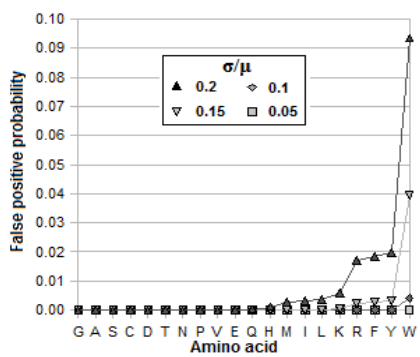


**Figure 4**. False positive probabilities for each of the standard 20 AAs. AA volumes are normally distributed with mean μ as given in Table 2. Probabilities are calculated for four values of standard deviation σ relative to μ.

Figure 5 shows the two types of errors averaged over all 20 amino acids weighted by their frequency of occurrence in the human proteome (Uniprot Id UP000005640; 11326153 residues, including 838 ambiguous residues excluded from the computation; see Figure S-2 in Supplementary File 1 for the frequency distribution.)

The error probabilities vary from negligible to moderate. The worst case is W, the largest volume AA, with a 17% false negative probability for σ = 0.2μ, the corresponding weighted error over the proteome is 0.095. These results suggest that determining whether the tRNA is charged or not is robust against wide variations in analyte

volume.

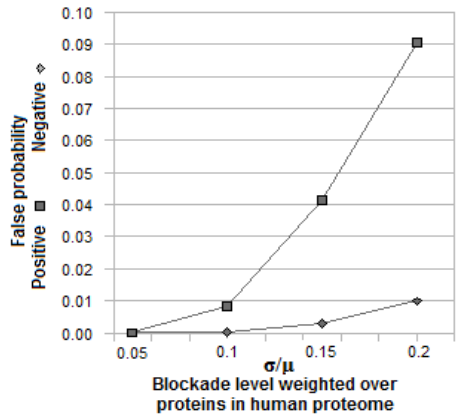


**Figure 5**. Weighted false negative and false positive probabilities for four values of standard deviation σ relative to mean μ. Probabilities are weighted by frequency of occurrence of 20 amino acids in human proteome (Uniprot Id UP000005640; 11326153 residues).

*Discriminating the smallest volume AA (Glycine) from baseline noise*

In Equation 6b, the blockade level due to the smallest amino acid, Glycine (G), must be distinguishable from the baseline noise $I_{RMS}$, usually defined as the rms (root-mean-square) value of the noise around the quiescent pore current I, which is the pore current in the absence of an analyte. Define the signal-to-noise ratio (SNR) as

$$SNR = I_{AN} / I_{RMS} \tag{14}$$

A sufficiently high SNR can ensure that the blockade current due to G can always be distinguished from $I_{RMS}$. In [28] the SNR for different kinds of pores, biological and synthetic, is given. It ranges from 4 to nearly 40, which is adequate for the present purpose. Various methods of increasing the SNR for both kinds of pores are discussed in [28]. In particular slowing down the analyte as it translocates through the pore can lead to a significant increase in SNR by reducing the bandwidth required for detection, and at the same time decreases high frequency noise. Figure 6 shows the probability that a measured blockade for Glycine is $< Th_{noise} = 2\ I_{RMS}$ for different values of $\sigma_{AA=G}$.

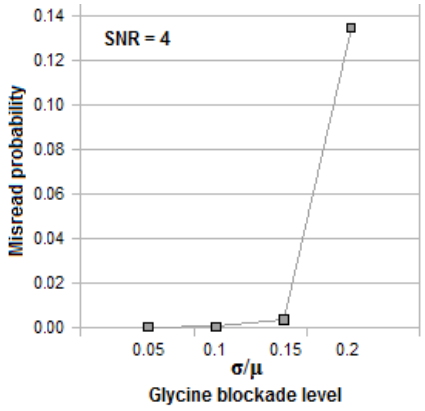


**Figure 6**. Probability of 2*baseline RMS noise ($I_{RMS}$) in nanopore current being mistaken for measured blockade due to smallest volume AA (G) with a signal-to-noise ratio (SNR) of 4 for different values of standard deviation σ relative to mean μ of volume of G.

## 5. Discussion

There is an extensive literature on practical issues affecting nanopores in polymer sequencing, and a nascent one on

the fabrication of nanocontainers. The following discussion is limited to the present context.

1) Stage 2 in Section 2 requires the identification of an AMP-level blockade when a tRNA has been charged. In most nanopore-based sequencing studies the speed with which an analyte translocates through the pore usually exceeds the bandwidth of the available detector. Various methods to control translocation speed are reviewed in [29,30]. The approach most useful in the present case involves the use of organic salts [31,32] or an RTIL (room temperature ionic liquid) [33], both of which are aimed at increasing solution viscosity; in [33] the use of RTIL as the electrolyte reduces the translocation speed of DNA by two orders of magnitude.

2) An analyte has to be drawn into the pore from the *cis* chamber, one way to do this is discussed in [34]. A second way is to reduce the bias in favor of positively or negatively charged AAs by reversing the applied voltage and repeating the observation. A much better solution is to use hydraulic pressure [20], which can guarantee delivery regardless of the electrical charge carried by the analyte. (Its use in confining reactants in a cavity has been discussed above (Figure 1c).)

3) In Stage 2 bunching of multiple AMP/ATP or free AA molecules can occur during translocation, leading to a missed read. In [18] a missed read is simply rejected. Here rejection is not an option, the solution lies in performing multiple measurements after reversing the voltage each time. This reduces the combined probability to zero.

4) In comparison with purely nanopore-based methods that use only electrical readout, this method involves non-trivial amounts of chemicals, such as multiple tRNAs, the associated AARSs, ATP, etc. However in many cases the reactants can be reused. For example, in Stage 2 the AARS enzyme can be saved from the *cis* chamber of the e-cell and accumulated for use later. Similarly the ATP molecules in *trans* can be reused.

5) The use of multiple passes (numbering on average 10 and a maximum of 20) to determine if a tRNA is charged increases the processing time considerably. However, pipelining can be used to decrease the total processing time over multiple amino acids when sequencing peptides or proteins. With N ($20 \leq N \leq 61$) copies of the amino acid and an array of N reservoir-cavities (Stage 1) and N e-cells (Stage 2) the number of passes can be reduced to one.

## 6. Conclusions

The proposed method has some significant conceptual and practical advantages. First, by shifting the bulk of the burden of AA identification to tRNAs and their cognate AARSs, it benefits from the accuracy and reliability that accompany superspecificity. Second, the absence of labeling [35] makes the method much simpler. Third, while nanopores play an important role in Stage 2 in the proposed approach that role is not a central one. Such properties as pore length, electro-osmotic effects, hydrophobicity, and others that affect conventional nanopore-based sequencing [4] are not important. The only major consideration is that the pore diameter be sufficiently small that a tRNA (charged or uncharged) or AARS cannot enter the pore.

In conclusion amino acids can be identified accurately without labeling by mimicking aspects of the cell's mRNA-to-protein translation process while retaining the latter's low error rate in charging a tRNA with a cognate amino acid. By making amino acid identification a predominantly chemical rather than physical process, limitations such as instrumentation precision and noise no longer have the importance they do in methods based on electrical, optical, or mass measurements. The proposed method can therefore be used as the core of a peptide/protein sequencing engine. As the individual steps are fairly simple and systematic the process can be automated to minimize or perhaps eliminate human intervention.

**Supplementary Information**. A Supplementary File contains details of the simulation, results covering a fuller range of parameters in an expanded version of Table 3, and other supporting data.

## References


[1] R. J. Simpson. *Proteins and Proteomics: A Laboratory Manual*, CSHL Press, 2008.

[2] E. de Hoffmann and V. Stroobant. *Mass Spectrometry: Principles and Applications*, 3rd edn., Wiley, 2007.

[3] L. Restrepo-Pérez, C. Joo, and C. Dekker. "Paving the way to single-molecule protein sequencing". *Nature Nanotech.*, **13**, 786-796, 2018. doi: 10.1038/s41565-018-0236-6

[4] W. Timp and G. Timp. "Beyond mass spectrometry, the next step in proteomics". *Sci. Adv.*, 2020, **6**, eaax8978.

[5] N. Callahan, J. Tullman, Z. Kelman, and J. Marino. "Strategies for development of a next-generation protein sequencing platform". *Trends Biochem. Sci*. 2019. doi:10.1016/j.tibs.2019.09.005

[6] F. Piguet, H. Ouldali, M. Pastoriza-Gallego, P Manivet, J Pelta, and A. Oukhaled. "Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore". *Nature Communs.*, 2018, **9**, 966. doi: 10.1038/s41467-018-03418-2

[7] H. Ouldali, K. Sarthak, T. Ensslen, F. Piguet, P. Manivet, J. Pelta, J. C. Behrends, A. Aksimentiev, and A. Oukhaled. "Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore". *Nature Biotech.*, December 16, 2019.

[8] J. Bush, W. Maulbetsch, M. Lepoitevin, B. Wiener, M. M. Skanata, W. Moon, C. Pruitt, and D. Stein. "The nanopore mass spectrometer". *Rev. Sci. Instruments*, 2017, **88**, 113307. doi: 10.1063/1.4986043
[9] N. Drachman, M. Lepoitevin, B. Wiener, H. Szapary, O. Isik, an D. Stein. "Towards single molecule protein sequencing by nanopore mass spectrometry". APS March Meeting 2019, abstract id S55.013.
[10] G. Bergtrom, *Cell and Molecular Biology, 3rd edn*. University of Wisconsin, Milwaukee, 2018.
[11] M. Ibba and D. Soll. "The renaissance of aminoacyl-tRNA synthesis". *EMBO Reports* **2**, 382-387, 2001.
[12] D. Goodsell. "Aminoacyl-tRNA Synthetases". *PDB Molecule of the Month*, April 2001. doi:10.2210/rcsb_pdb/mom_2001_4
[13] D Pedone 1, M Langecker, A M Münzer, R Wei, R D Nagel, U Rant. "Fabrication and electrical characterization of a pore-cavity-pore device". *J Phys Condens Matter*, 2010, 22, 454115. doi: 10.1088/0953-8984/22/45/454115
[14] D. Pedone, M. Langecker, G. Abstreiter, and U. Rant. "A pore-cavity-pore device to trap and investigate single nanoparticles and DNA molecules in a femtoliter compartment: confined diffusion and narrow escape". *Nano Lett*. 2011, 11, 1561–1567.
[15] X. Liu, M. Skanata, and D. Stein. "Entropic cages for trapping DNA near a nanopore". *Nature Communs.,* 2015, **6**, 6222.
[16] H. Caoa, Z. Yu, J. Wang, J. O. Tegenfeldt, R. H. Austin, E. Chen, W. Wu and S. Y. Chou. "Fabrication of 10 nm enclosed nanofluidic channels". *Appl. Phys. Lett*., 2002, **81**, 174-176. doi: 10.1063/1.1489102
[17] A. P. Ivanov, P. Actis., P. Jönsson, D. Klenerman, Y. Korchev, and J. B. Edel. "On-demand delivery of single DNA molecules using nanopipets". *ACS Nano*, 2015, **9**, 3587–3595.
[18] M. Rahman, M. A. Stott, M. Harrington, Y. Li, M. J. N. Sampad, L. Lancaster, T. D. Yuzvinsky, H. F. Noller, A. R. Hawkins, and H. Schmidt. "On demand delivery and analysis of single molecules on a programmable nanopore optofluidic device". *Nature Communs*., 2019, **10**, 3712. doi.org/10.1038/s41467-019-11723-7
[19] D. L. Nelson and M. M. Cox. *Lehninger's Principles of Biochemistry*, 4th edn., W H Freeman, 2005.
[20] B. Lu, D. P. Hoogerheide, Q. Zhao, H. Zhang, Z. Tang, D. Yu, and J. A. Golovchenko. "Pressure-controlled motion of single polymers through solid-state nanopores." *Nano Lett*., 2013, **13**, 3048–3052.
[21] M. Kolmogorov, E. Kennedy, Z. Dong, G. Timp, and P. Pevzner. "Single-molecule protein identification by sub-nanopore sensors". *PLoS Comp. Biol*., 2017. doi: 10.1371/journal.pcbi.1005356
[22] S. J. Perkins. "Protein volumes and hydration effects". *Eur. J. Biochem*., **157**, 169-180, 1986.
[23] M. Zwolak and M. Di Ventra. "Physical approaches to DNA sequencing and detection". *Rev. Mod. Phys*., 2008, **80**, 141-165.
[24] J. E. Reiner, A. Balijepalli, J. W. F. Robertson, B. S. Drown, D. L. Burden, and J. J. Kasianowicz. "The effects of diffusion on an exonuclease/nanopore-based DNA sequencing engine". *J. Chem. Phys*., 2012, **137**, 214903. doi: 10.1063/1.4766363
[25] K. T. Brady and J. E. Reiner, "Improving the prospects of cleavage-based nanopore sequencing engines," *J. Chem. Phys*., 2015, 143, 074904.
[26] G. Sampath. "DNA sequencing with stacked nanopores and exonuclease: a simulation-based analysis". *Electrophoresis*, 2016, **37**, 2429-2434. doi: 10.1002/elps.201600049
[27] J. Clarke, H. C. Wu, L. Jayasinghe, A. Patel, S. Reid, and H. Bayley. "Continuous base identification for single-molecule nanopore DNA sequencing". *Nature Nanotech*., 2009, **4**, 265–70.
[28] A. Fragasso, S. Schmid, and C. Dekker. "Comparing current noise in biological and solid-state nanopores". *bioRxiv.org*, 2019. doi: 10.1101/866384
[29] U. F. Keyser. "Controlling molecular transport through nanopores". *J. R. Soc. Interface* **8**, 1369-1378, 2011.
[30] S. Carson and M. Wanunu. "Challenges in DNA motion control and sequence readout using nanopore devices". *Nanotechnology* , 2015, **26**, 074004. doi:10.1088/0957-4484/26/7/074004
[31] R. Samanthi, S. de Zoysa, D. A. Jayawardhana, Q. Zhao, D. Wang, D. W. Armstrong, and X. Guan. "Slowing DNA translocation through nanopores using a solution containing organic salts". *J. Phys. Chem. B*, 2009, **113**, 13332–13336.
[32] S. W. Kowalczyk, D. B. Wells, A. Aksimentiev, and C. Dekker. "Slowing down DNA translocation through a nanopore in lithium chloride". *Nano Lett*. 2012, **12**, 1038-1044. doi:1 0.1021/nl204273h
[33] J. Feng, K. Liu, R. D. Bulushev, S. Khlybov, D. Dumcenco, A. Kis, and A. Radenovic. "Identification of single nucleotides in $MoS_2$ nanopores". *Nature Nanotech*., 2015. doi: 10.1038/nnano.2015.219
[34] K. J. Freedman, L. M. Otto, A. P. Ivanov, A. Barik, S.-H. Oh, and J. B. Edel. "Nanopore sensing at ultra-low concentrations using single-molecule dielectrophoretic trapping". *Nature Commun*. 2016, **7**, 10217. doi: 10.1038/ncomms10217
[35] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha. "Advances in single-molecule fluorescence methods for molecular biology." *Annu Rev Biochem*., 2008, **77**, 51–76.