

Amino acid identification via tRNA charging, deacylation, and current blockade in a nanopore

G. Sampath

P. B. 7849, J. P. Nagar P. O., Bengaluru-560078, India

Abstract. A label-free four-step procedure to identify amino acids (AAs) is described. In Step 1 molecules of AA, a tRNA, the related tRNA synthetase (AARS), and ATP are confined in a cavity to enable charging of tRNA. In Step 2 the tRNA, charged or uncharged, is separated from the other reactants (ATP, AARS, and possibly AMP and free AAs). In Step 3 the separated tRNA is subjected to non-enzymatic deacylation to dissociate the AA if tRNA is charged. In Step 4 the products are transferred to an electrolytic cell with a nanopore, where a current blockade occurs if and only if there is a dissociated AA. If a blockade is observed AA is immediately known, identification is unambiguous because of tRNA superspecificity. The exact blockade size need not be known (for any AA) in Step 4 so there is no need to distinguish among different AAs. This is unlike other nanopore-based methods, which are crucially dependent on precise blockade level measurements. The procedure is done in parallel with N ($20 \leq N \leq 61$) copies of AA, N reservoir-cavities, and N pairs of e-cells each with a different tRNA, related AARS, and ATP. At least one of the tRNAs gets charged. Assuming no charging errors, if N_1 tRNAs get charged in Step 1 and at least one of N_2 ($\leq N_1$) charged tRNAs is deacylated in Step 3, then with $N_1 N_2$ tRNA molecules entering Step 1 AA can be identified with probability 1.

Keywords: AARS; deacylation; label-free; nanopore; current blockade; protein sequencing

1. Introduction

Current protein/peptide sequencing is mostly based on one of two bulk methods: 1) Edman degradation of the N-terminal residue and spectroscopic identification of the latter [1]; and 2) mass spectrometry (MS) [2]. More recently single molecule methods based on nanopores have emerged as a possible alternative. In nanopore-based polymer sequencing a single DNA, RNA, or protein molecule translocates electrophoretically through a nano-sized pore in an electrolytic cell (e-cell) and causes current blockades that are used to identify the sequence of monomers [3,4]. The current (c. 2019) state of protein sequencing research is reviewed in [5].

Related work focuses on recognizing single residues within a peptide or free amino acids; see for example [6] and [7]. The latter reports that 13 of the 20 standard amino acids (AAs) can be identified solely from their volume as they translocate through the pore. In [8,9] nanopores are combined with standard MS methods. In a somewhat different approach a label-free method has been proposed [10] for the identification of single amino acids based on the superspecificity property of tRNAs [11]. Superspecificity ensures that a tRNA is almost always charged with its cognate amino acid by a related amino-acyl tRNA synthetase (AARS), with adenosine triphosphate (ATP) providing the energy. Charging is highly accurate, with *in vivo* error rates around 1 in 250 [12]. Most recently naphthalene-based derivatives of 9 of the 20 amino acids have been used as surrogates to increase their variability and thereby allow their identification from current blockades [13].

1.1 The present work.

Here a modified version of [10] is proposed and analyzed. In [10] molecules of a specific AA, a tRNA and its cognate AARS, and adenosine triphosphate (ATP) from a reservoir are restricted to a micro-sized cavity to counter the effects of diffusion. 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. The occurrence of an AMP-sized blockade means that a tRNA molecule has been charged with AA, this identifies AA from the tRNA. The procedure is done in parallel with N ($20 \leq N \leq 61$) copies of AA, N cavities, and N e-cells, each with ATP and a different tRNA and its cognate AARS. At least one of the tRNAs will get charged with AA. This approach requires being able to discriminate AMP blockades from those due to any of the AAs or ATP.

In the modified version given here the tRNA molecules (charged or uncharged) in the cavity are separated from the other reactants and then deacylated to dissociate the AA from a charged tRNA. The products of the deacylation step are transferred to an e-cell where the dissociated AA causes a current blockade. AA is then known from the identity of the tRNA. If the tRNA is not charged there can be no blockade and AA is not identified. As before the procedure is done in parallel with N copies of AA, cavities, e-cells, and reactants; one of the tRNAs gets charged and AA is identified. This approach considerably simplifies the identification procedure because there is no need to distinguish among blockades due to different analytes (AA, ATP, AMP). Just the occurrence of a blockade distinguishable from noise is sufficient to identify an amino acid.

2. A four-step amino acid identification procedure

There are four steps in the modified procedure:

- 1) charging of tRNA with AA in a confined space (same as Step 1 in [10]);
- 2) separation of tRNA (charged or uncharged) from other reactants;
- 3) deacylation of charged tRNA; and
- 4) identification of AA from current blockade in a nanopore and identity of tRNA (or none if tRNA is uncharged).

Figure 1 shows the four stages, which correspond to (a), (b), (c, e), and (d, f).

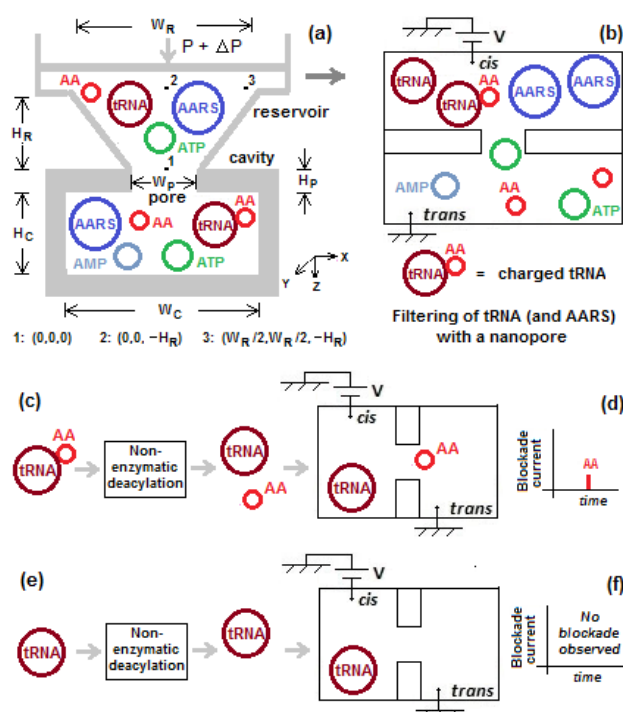


Fig. 1 (a) Reservoir-cavity structure with hydraulic pressure to confine reactants to cavity (1, 2, 3 correspond to three particle origins used in simulation); (b) Nanopore in electrolytic cell (e-cell) filters charged or uncharged tRNA (and AARS) from other reactants; (c, d) Filtered tRNA (charged) deacylated non-enzymatically (see text); products (deacylated tRNA, free AA) transferred to *cis* chamber of second e-cell, free AA causes blockade; (e, f) Filtered tRNA (uncharged), deacylation has no effect; tRNA enters second e-cell, no free AA, no blockade. (AARS filtered out, not shown in c, d, e, and f.) Figure not to scale.

Step 1: Charging of tRNA

This step is identical to Step 1 in [10]. AA, tRNA, AARS, and ATP are confined to a small cavity to avoid the dispersive effects of diffusion. A modified version of the 'cage' in [14] is used. AA, tRNA, AARS, and ATP pass from a reservoir into a cavity where a cognate tRNA can get charged with AA. The reactants remain inside the cavity under hydraulic pressure. The latter has been used to study the translocation properties of DNA in a nanopore [15]. As it is independent of electrical charge it works equally well with all the analyte types involved.

Step 2: Separation of tRNA (charged or uncharged)

After a suitable incubation time in the cavity (to be determined experimentally) the products are largely tRNA, AA, AARS, ATP, and AMP. They enter Step 2, in which tRNA is separated from the other reactants. (Whether it is charged or not will be known only in Step 4.) Separation of tRNA can be done into the *cis* chamber of an e-cell with a nanopore of appropriate size, which can be determined from the analyte sizes given in Table 1 below. In Figure 1(b) tRNA and AARS are separated from AA, ATP, and AMP, with the latter translocating through the pore to *trans*. If desired AARS can be filtered out, although it may not be necessary.

Step 3: Deacylation of charged tRNA

The tRNA molecules that remain in the *cis* chamber of the first e-cell (Figure 1b) are deacylated non-enzymatically (using a reagent like NaOH or by pH control [16]). The products of this step are either tRNA and the dissociated AA (if tRNA was charged) or only tRNA.

Step 4: Identification of AA from current blockade

The products of the deacylation step are transferred to the *cis* chamber of a second e-cell (Figure 1d and 1f). If the tRNA is charged the dissociated AA can pass into *trans* but not tRNA. In doing so it causes a current blockade. If tRNA is not charged there is no current blockade because there is no free AA. In the former case AA is immediately known from the identity of the tRNA. The only requirement is that the blockade level due to the smallest volume amino acid, Glycine (G), be distinguishable from baseline noise; this is discussed later. This development leads to the following determination:

$$\text{Current blockade in 2}^{\text{nd}} \text{ e-cell} \rightarrow \text{tRNA is deacylated} \rightarrow \text{AA known from identity of tRNA} \quad (1)$$

Conversely

$$\text{No current blockade in 2}^{\text{nd}} \text{ e-cell} \rightarrow \text{tRNA not deacylated} \rightarrow \text{tRNA not cognate to AA} \quad (2)$$

The procedure is done in parallel with N ($20 \leq N \leq 61$) copies of AA, N reservoir-cavities, and N pairs of e-cells each with a different tRNA, related AARS, and ATP. As there is at least one tRNA cognate to AA one of the N tRNAs gets charged.

As in [10] it is assumed that the size of the current blockade I_B in the second e-cell is roughly proportional to V_{AN} , the volume excluded when an analyte AN passes through the pore [7]:

$$I_B \sim V_{AN} \quad (3)$$

In what follows V_{AN} is used as a proxy for I_B .

As it is only necessary to observe a blockade due to an AA that has been dissociated from a charged tRNA the observed exclusion volume due to AA will satisfy

$$V_G < V_{AN} < V_W \quad (4)$$

where V_G is the volume of the smallest amino acid Glycine (G), and V_W that of the largest amino acid Tryptophan (W). See Table 1 below (this is a subset of Table 2 in [10]).

Table 1. Volumes of the standard 20 amino acids from [17], given as mean values in \AA^3 or $\times 10^{-3} \text{ nm}^3$.

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

3. Analysis of the amino acid identification procedure

The following assumptions are made:

- 1) Superspecificity holds *in vitro*. The literature on *in vitro* experiments involving AARSs supports this assumption.
- 2) With a suitable number of reactant molecules (AARS, ATP, AA) and favorable environmental conditions (temperature, buffer, stabilizer, etc.) at least some of the tRNA molecules get charged and the charged state lasts long enough that filtration (Step 2) and deacylation (Step 3) can be performed without any interference from environmental agents. The results of the extensive *in vitro* experiments on lysyl-tRNA described in [18] (with product lifetimes in the tens of minutes) make this a reasonable assumption.
- 3) Environmental agents (for example, buffer) and other chemicals (such as non-enzymatic reagents for deacylation) do not have a confounding effect on the observation of blockades due to any of the 20 amino acids. A perusal of the literature on nanopore-based methods suggests that buffering and non-enzymatic agents do not cause observable

blockades.

Conditions for correct identification of an amino acid

The following conditions have to be satisfied in order for the procedure in Section 2 to correctly identify an amino acid:

- 1) The reactants necessary for charging of a tRNA must be in close proximity until successful charging occurs.
- 2) Perfect separation of tRNAs occurs in Step 1.
- 3) Deacylation of a charged tRNA always occurs.
- 4) Current blockades due to any AA can always be distinguished from baseline noise.

It is now argued that the above conditions are satisfied in the proposed procedure.

Condition 1. Reactants in the cavity can be brought together with a directed force that can overcome the effects of diffusion. In Step 1 this is achieved with a hydraulic force. While diffusion is isotropic the directed hydraulic force biases the movement of particles in the z direction (see Figure 1a) thereby ensuring that they accumulate at the bottom of the cavity. Such confinement is further reinforced if the cavity has a tapered geometry. In [10] the efficacy of this arrangement was studied by simulation. The results there show that the particle can remain indefinitely near the bottom of the cavity, which is just what is desired. This conclusion is further strengthened if one were to consider the motive force when it is purely diffusive (no hydraulic pressure). Simulation data in [10] show the particle to be still diffusing in the reservoir at the end of a run, showing that hydraulic pressure is the key to successful confinement.

Condition 2. The dimensions of the molecules coming out of Step 1, namely tRNA, AARS, AA, AMP, and ATP, can be approximated with the axes of enclosing ellipsoids, which can be calculated from atomic data. The results are shown in Table 2 (which is a subset of Table 1 in [10]). Clear dividing lines that separate tRNA from AARS on one side and from AMP, ATP, and any of the AAs on the other, can be seen. Thus separation of tRNA, charged or not, from the other reactants can be easily done with a nanopore of an appropriate size obtained from the data in Table 2.

Table 2. Dimensions of reactants involved in tRNA charging. Amino acids represented by smallest (Glycine) and largest (Tryptophan).

Reactant	Axes of ellipsoid enclosing reactant			Reactant	Axes of ellipsoid enclosing reactant		
	Major (nm)	Minor (nm)	Minor (nm)		Major (nm)	Minor (nm)	Minor (nm)
AMP ¹	1.67	0.95	0.4	Tryptophan (AA) ⁴	1.10	0.7	0.34
ATP ²	2.34	1.09	0.44	Phe tRNA ⁵	10.56	6.54	3.64
Glycine (AA) ³	0.56	0.35	0.21	Ala AARS ⁶	13.94	9.41	8.46

^{1,2,3,4,6} Atomic coordinate data from files.rcsb.org/ligands/view/AMP_ideal.sdf, files.rcsb.org/ligands/view/ATP_ideal.sdf, www.rcsb.org/ligand/GLY, files.rcsb.org/ligands/view/TRP_ideal.sdf, www.rcsb.org/structure/1YFS. (The last is for Ala AARS.)

⁵ Atomic coordinate data from [19].

Condition 3. In the living cell deacylation is done by an enzyme to correct charging errors [11]. Thus if a tRNA is misacylated an editing AARS deacylates the tRNA and attaches the correct AA to it. A table of Class I and Class II AARSs and their AA editing targets is given in [20]. While enzymatic deacylation can be used in Step 3, it adds unnecessary complexity. Non-enzymatic deacylation is much simpler. The latter may be done with NaOH or by controlling the pH level of the solution. See [16] for a description of its use in the deacylation of Cysteine-tRNA.

Condition 4.

Consider Equation 4. The blockade V_G of the smallest amino acid, Glycine (G), must be distinguishable from the volume V_{RMS} corresponding to the baseline noise I_{RMS} , which is the rms (root-mean-square) value of the noise around the quiescent pore current I , the pore current in the absence of an analyte. Writing the signal-to-noise ratio (SNR) as

$$SNR = V_{AN} / V_{RMS} \quad (5)$$

a sufficiently high SNR can ensure that the blockade current due to G can always be distinguished from I_{RMS} . In [21] 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. See [10] for a discussion.

Probability of correct identification

The probability that AA is correctly identified can be estimated by considering the number of molecules of each reactant in the cavity in Step 1 and the number of charged tRNAs that enter Step 3 for deacylation. Assuming no charging errors, if N_1 tRNAs get charged in the first step and at least one of N_2 ($\leq N_1$) charged tRNAs is deacylated, then with $N_1 N_2$ tRNA molecules in Step 1, AA can be identified with probability 1. To get an idea of the quantities of reactants required, consider conservative charging and deacylation rates of 1%, corresponding to $N_1 = N_2 = 100$. This would require $N_1 N_2 = 10000$ molecules of tRNA entering Step 1. This number corresponds to about 15 zeptomoles. With rates of 10% $N_1 N_2 = 100$, which is about 160 yoctomoles.

4. Conclusions

This report shows that amino acids can be identified accurately without labeling by mimicking aspects of the cell's mRNA-to-protein translation process [22] while retaining the latter's low error rate in charging a tRNA with a cognate amino acid. This is achieved in two ways: 1) 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; and 2) By deacylating a charged tRNA, identification is reduced to observing a current blockade without having to know its exact value. The mere occurrence of a blockade is sufficient to identify AA from the identity of the tRNA.

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.*, 2018, **13**, 786-796. 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 Commun.*, 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.*, 2020, **38**, 176–181.
- [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, and D. Stein. "Towards single molecule protein sequencing by nanopore mass spectrometry". APS March Meeting 2019, abstract id S55.013.
- [10] G. Sampath. "Label-free amino acid identification based on tRNA charging in a cavity and AMP blockade in a nanopore". *bioRxiv.org*, June 26, 2020. doi: 10.1101/2020.06.25.170803
- [11] M. Ibba and D. Soll. "The renaissance of aminoacyl-tRNA synthesis". *EMBO Reports*, 2001, **2**, 382-387.
- [12] D. Goodsell. "Aminoacyl-tRNA Synthetases". *PDB Molecule of the Month*, April 2001. doi:10.2210/rcsb_pdb/mom_2001_4
- [13] X. Wei, D. Ma, L. Jing, L. Y. Wang, X. Wang, Z. Zhang, B. J. Lenhart, Y. Yin, Q. Wang, and C. Liu. "Enabling nanopore technology for sensing individual amino acids by a derivatization strategy". *J. Mater. Chem. B*, 2020. doi: 10.1039/D0TB00895
- [14] X. Liu, M. Skanata, and D. Stein. "Entropic cages for trapping DNA near a nanopore". *Nature Commun.*, 2015, **6**, 6222.
- [15] 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.
- [16] H. Jakubowski. "Editing function of Escherichia coli cysteinyl-tRNA synthetase: cyclization of cysteine to cysteine thiolactone". *Nucl. Acids Res.*, 1994, **22**, 1155-1160.
- [17] S. J. Perkins. "Protein volumes and hydration effects". *Eur. J. Biochem.*, 1986, **157**, 169-180.
- [18] T. Takita. "The structure and functions of the lysyl-tRNA synthetase of bacillus stearothermophilus". Dissertation, Kyoto University, March 1996. doi: 10.11501/3110616
- [19] G. J. Quigley, N. C. Seeman, A. H. -J. Wang, F. L. Suddath, and A. Rich, "Yeast phenylalanine transfer RNA: atomic coordinates and torsion angles". *Nucl. Acids Res.*, 1975, **2**, 2329-2341.
- [20] M. A. R. Gomez and M. Ibba. "Aminoacyl-tRNA Synthetases". *RNA* (online), April 17, 2020.
- [21] A. Fragasso, S. Schmid, and C. Dekker. "Comparing current noise in biological and solid-state nanopores". *bioRxiv.org*, 2019. doi: 10.1101/866384
- [22] G. Bergtrom, *Cell and Molecular Biology*, 3rd edn. University of Wisconsin, Milwaukee, 2018.

Email: sampath_2068@yahoo.com