Label-free amino acid identification for *de novo* protein sequencing via tRNA charging and current blockade in a nanopore

G. Sampath

Abstract
A label-free procedure to identify single amino acids (AAs) for *de novo* protein sequencing is developed in theory and simulated in part. A terminal AA cleaved from a protein/peptide, a tRNA, its cognate amino-acyl tRNA synthetase (AARS), and adenosine triphosphate (ATP) are brought together in a container where tRNA, if cognate, gets charged with AA and adenosine monophosphate (AMP) is released. The tRNAs (charged or not) are filtered out, deacylated, and transferred to an electrolytic cell (e-cell) with a nanopore. If a blockade occurs it can only be due to AA; this immediately identifies AA, the error rate is < 1/350 because of tRNA superspecificity. This procedure can be executed in parallel with 20 or more molecules of AA and 20 different tRNAs to identify AA in one cycle. (It can also be done in series if the sample size is small, with only a small number of molecules available.) In a second version the released AMP translocates with ATP and free AA into the *cis* chamber of an e-cell. If tRNA is charged an AMP blockade is observed, this identifies AA. This version is easier to implement and parallelize (or serialize, if only a few molecules are available). In the first method it suffices to distinguish a blockade from noise; in the second an AMP blockade must be differentiated from one due to ATP or any of the 20 AAs. Even so the error is < 10% when weighted over the human proteome. Both are *de novo* methods; no prior information about the protein is needed. Unlike most other nanopore-based methods they do not need to discriminate among the 20 proteinogenic AAs when measuring/observing blockades.

Keywords single molecule; nanopore; label-free; AARS; protein sequencing; superspecificity

1. Introduction

Sequencing of a peptide or protein is commonly based on Edman degradation (ED) plus chromatography [1,2] or on mass spectrometry (MS) [3]; both are bulk methods. Protein identification, in contrast, may be done without using any sequence knowledge or from a partial sequence. In the former case chemical methods may be used with affinity-based reagents to identify a target protein [4]. Physical methods include the use of multiple physical properties [5] or carrier molecules [6], they require knowledge of the target. With partial sequences a subsequence may be compared with a protein sequence database to uniquely identify a protein in the proteome [7]. For example, in [8] the volumes of short subsequences are sufficient to identify 90% of the proteins in *Helicobacter pylori*.

The present work focuses on full *de novo* sequencing that does not use prior knowledge of the protein/peptide in any form.

For a recent review of protein sequencing research see [9]. In [10] a protein molecule is broken into short peptides and the carboxyl end of each peptide immobilized on a surface. Optical tags attached to selected residue types, the terminal residues are removed in succession with ED and the resulting change in the fluorescence level used to identify tagged residues. This yields a partial sequence that is sufficient to identify the parent protein in a sequence database; full *de novo* sequencing is yet to be realized with this method. A theoretical model is described in [11] where the binding times of optically tagged sensor N-terminal AA binders (NAABs) [12] to the N-terminal residue of the peptide are calculated and used to identify N-terminal residues. This method has the potential for full sequencing.

Nanopores provide an electrically-based single molecule approach [13]. Nanopore protein sequencing [14-16] is still in its infancy. A variety of methods are known, some in theory [17,18], others in practice [19,20]. In [21] 13 of the 20 proteinogenic AAs are identified from the volume excluded as the AA translocates through a nanopore. In [22] naphthalene-based derivatives of 9 of the 20 AAs are used as surrogates to allow easier identification from nanopore blockades.

1.1 The present work

Here a method for AA identification aimed at *de novo* protein sequencing that can be used in the bulk or at the single molecule level is proposed and simulated in part. It combines chemical-based methods with nanopore-based sensing At its center is the 'superspecificity' [23] property of transfer RNAs (tRNAs), on which translation of mRNA to protein in living cells is based [24]: For every AA there is a tRNA that binds to that AA and to no other;

the *in vitro* error rate is about 1 in 350 [25]. Binding is accompanied by the release of adenosine monophosphate (AMP).

Two versions are described. In the first, AA, a tRNA and its cognate AARS, and adenosine triphosphate (ATP) are confined in a cavity to enable charging of a tRNA molecule with its cognate AA. The tRNA is filtered out and deacylated. The products are transferred to an electrolytic cell (e-cell) with a nanopore where the free AA causes a current blockade. This unambiguously identifies AA. When executed in parallel on 20 or more copies of AA each with a different tRNA, AARS, and ATP, AA is identified in a single cycle. (If only a few molecules are available, as would be the case if the sample were to be obtained from single cells, the procedure can be applied in series with a different tRNA and AARS in each step.) By feeding terminal AAs cleaved from a peptide/protein to this parallel (or series) processor, de novo sequencing can be done. In the second version the AMP released by charging enters the *cis* chamber of an e-cell (along with free AA and ATP) and translocates through the pore to cause a blockade; this identifies AA. This version is easier to implement and parallelize (or serialize). Although the method is a little less accurate, the error is still less than 10% when the results are weighted over the human proteome.

2. A four-step procedure to identify an amino acid

The method is centered on the charging of a tRNA molecule with a cognate AA. It may be done in the bulk or on a few molecules. The following equations summarize the charging process [24]:

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

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

In an aqueous solution the pyrophosphate (PPi) is unstable and breaks up:

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

The procedure consists of four steps, corresponding to (a,b), (b,c), (c,d), and (e/f) or (g/h) in Figure 1, where the last step distinguishes between two cases: charged and uncharged tRNA.

Step 1: *Charging of tRNA*

Molecules of AA, tRNA, AARS, and ATP are brought together in solution (Figure 1a). The container may be a conventional test-tube, or a nano-size container if only trace amounts of the sample are available. In the latter case the reactants must remain close proximity without diffusing away. This can be achieved with hydraulic pressure [26], which works with all the analyte types involved, charged or not. Simulations of a simple hydrodynamic model based on Poiseuille flow shows that the reactants are confined to the 'bottom' of the container; see below. The details of simulation are given in the Supplement.

Table S-1. Dimensions of reactants involved in tRNA charging. Amino acids are represented by smallest (Glycine) and largest (Tryptophan), tRNAs by Phe-tRNA, and AARSs by Ala-AARS.

Space or reactant	Dimensions of box (or ellipsoid axes)			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				

[1,2,3,4,6]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 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". *Nucleic Acids Research*, **2**, 2329-2341, 1975.

Step 2: *Separation of tRNA (charged or uncharged)*

After a suitable incubation time in the container (to be determined experimentally) the products (Figure 1b) may include charged cognate tRNA, uncharged tRNA, free AA, AARS, ATP, and AMP. The last will be present only if

tRNA is cognate tRNA at least one molecule has been charged). tRNA and AARS are separated from AA, ATP, and AMP (if any) (Figure 1c); next the tRNA is separated from AARS (Figure 1d). The filter size required is calculated from the axes of enclosing ellipsoids for each of the analytes involved. Table 1 above shows the reactant sizes and the pore sizes of filters to use (3 nm and 12 nm). Solid state nanopores with these pore sizes are available.

Step 3: *Deacylation of charged tRNA*
The tRNA, if charged, can be deacylated to release the bound AA, which can then be detected in the nanopore of an e-cell. Deacylation can be done enzymatically or without an enzyme. Non-enzymatic deacylation is much simpler. It may be based on the use of NaOH or by controlling the solvent's pH level [27]. The products of this step are either tRNA and the dissociated AA (if tRNA was charged) or only tRNA.

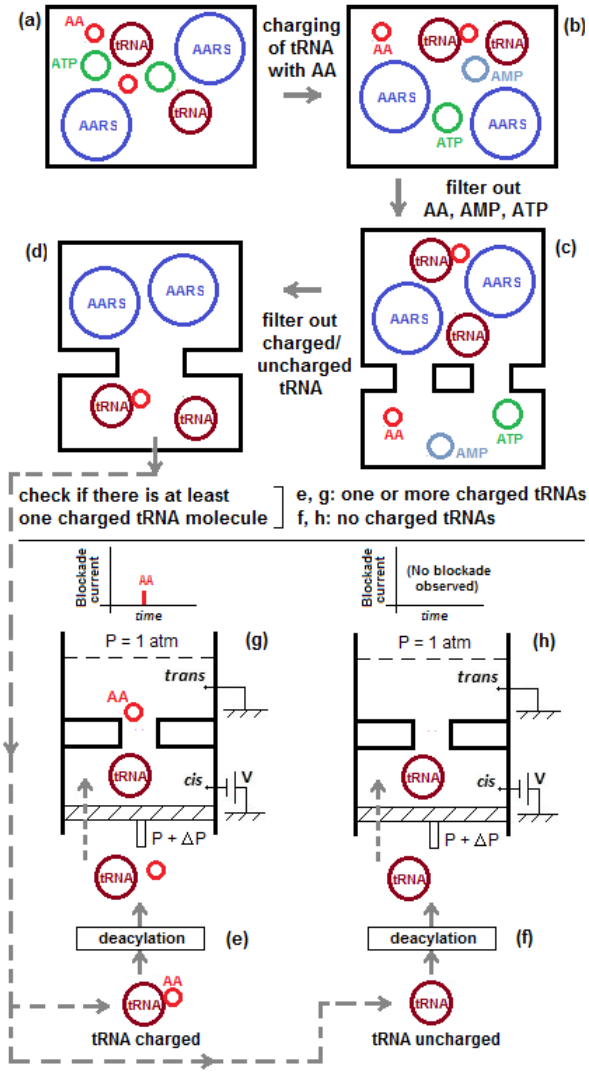

Fig. 1 Amino acid identification process. (a,b) Charging of $tRNA_i$ ($0 \leq i \leq N$), one of N ($20 \leq N \leq 61$) different tRNAs; (c,d) separation of $tRNA_i$, charged or not; (e) deacylation of charged $tRNA_i$; (f) deacylation of uncharged $tRNA_i$, no effect; products transferred to *cis* chamber of e-cell; (g) If $tRNA_i$ is charged free AA causes blockade → AA known from $tRNA_i$; (h) if $tRNA_i$ not charged, no blockade → AA not identified. Schematic, not to scale.

Step 4: *Identification of AA from current blockade*
The products of deacylation are transferred to the *cis* chamber of an e-cell (Figure 1d and 1f) with a 3 nm pore (see Table 1). If the tRNA is charged the dissociated AA (but not tRNA) can pass into *trans*, and cause a current blockade. If tRNA is not charged there is no free AA and no current blockade. 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.

These steps lead to the following simple determination:

Current blockade in e-cell → tRNA is deacylated → AA known from identity of tRNA (4)

Conversely

No current blockade in e-cell → tRNA not deacylated → tRNA not cognate to AA (5)

Assuming that the size of the current blockade I_B in the e-cell is roughly proportional to V_{AN}, the volume excluded when an analyte AN passes through the pore [20,21]

$$I_B \sim V_{AN} \tag{6}$$

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 satisfies

$$V_G \leq V_{AN} \leq V_W \tag{7}$$

where V_G is the volume of the smallest amino acid Glycine (G), and V_W that of the largest amino acid Tryptophan (W). From Table S-2 in the Supplement, $V_G = 59.9 \times 10^{-3}$ nm^3 and $V_W = 227.9 \times 10^{-3}$ nm^3.

Feasibility and accuracy: conditions for correct identification of an amino acid

The feasibility of the method described above can be judged from the feasibility of the parts: tRNA charging; AMP release; nanopore sensing of AA blockades; hydraulic pressure in a nanopore; non-enzymatic deacylation; and cleaving of a terminal residue from a protein. All of these are based on known and well-established laboratory techniques. The accuracy of the method can be judged based on a set of necessary conditions for correct identification of an AA.

The following assumptions are made:

1) Superspecificity holds *in vitro*. The considerable literature on *in vitro* experiments involving AARSs supports this assumption; for a recent brief review see [28].

2) With enough reactant molecules (AARS, ATP, AA) and favorable environmental conditions (temperature, buffer, etc.) at least some of the tRNA molecules get charged. The results of *in vitro* experiments on lysyl-tRNA described in [29], with product lifetimes in the tens of minutes, make this a reasonable assumption.

3) Buffers and other reagents (such as NaOH for deacylation) do not confound observation of blockades due to any of the 20 amino acids. This is supported by the extensive literature on blockade measurements in nanopores.

The following conditions have to be satisfied in order for the above procedure 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) tRNA molecules are perfectly separated from the other reactants.

3) Deacylation of a charged tRNA always occurs.

4) Current blockades due to any AA can always be distinguished from baseline noise.

Condition 1. With bulk samples, this condition is easily achieved. With only trace amounts of AA available, the available AA molecules must be confined in a sub-micron-sized cavity with ATP and a tRNA and its related AARS to circumvent the dispersive effects of diffusion. Confinement enables charging of tRNA with AA if cognate; it can be achieved by bringing together the above reactants in the cavity with a directed force that can overcome the effects of diffusion. Hydraulic force is appropriate [26]. Thus while diffusion is isotropic the directed hydraulic force biases the movement of particles toward the bottom of the cavity thus ensuring that they accumulate there. This process is simulated with Monte-Carlo techniques by assuming an analyte molecule to be a dimensionless particle. Figure 2 shows a schematic of the reservoir-cavity structure simulated. The procedure is similar to the procedures in [30,31]. See the Supplement for a description of the geometry, protocol, and computations involved.

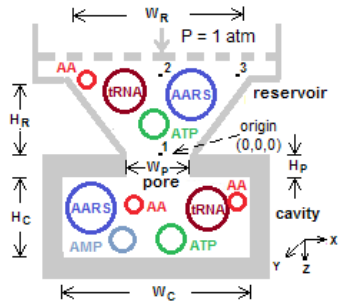

Figure 2. Schematic of reservoir-cavity structure showing confinement of reactants in cavity under hydraulic (atmospheric) pressure

To assess the efficacy of confinement the reservoir-cavity structure is simulated with a range of parameters (diffusion coefficient, hydraulic pressure level, origin of particle, and number of diffusion steps). The following measures are used: 1) Time of entry of particle into cavity; 2) Time to reach bottom of cavity; and 3) Dwell time at bottom of cavity. The simulation is stopped when the dwell time exceeds 10^6 continuous steps. In all runs the particle was found to always settle at the bottom of the cavity. Table 2 below shows the results. The results suggest 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.(Compare this with the results obtained in [32,33] with similar structures, the confinement times range from a few milliseconds [32] to tens of minutes [33]. The lower part of Table 2 shows the results for when the motive force is exclusively diffusive (no hydraulic pressure). The results show that the particle is still diffusing in the reservoir at the end of the run. Hydraulic pressure is thus the key to confinement of the reactants to the cavity.

Table 2 Simulation of dimensionless particle in reservoir-cavity structure under hydraulic pressure and diffusion

D	Origin	N_1 ($\times 10^{-6}$ s)	N_2 ($\times 10^{-6}$ s)	N_3 ($\times 10^{-6}$ s)
0.3	(0,0,0)	1.21	1.38	> 1.0
	(0,0,-H_R)	1.53	1.70	> 1.0
	(W_R/2, W_R/2, -H_R)	1.62	179	> 1.0
3.0	(0,0,0)	1.23	1.40	> 1.0
	(0,0,-H_R)	1.47	1.64	> 1.0
	(W_R/2, W_R/2, -H_R)	1.62	1.79	> 1.0
30.0	(0,0,0)	1.18	1.34	> 1.0
	(0,0,-H_R)	1.25	1.41	> 1.0
	(W_R/2, W_R/2, -H_R)	1.63	1.80	> 1.0

Pure diffusion					
Origin	N_3 [(1)] ($\times 10^8$)	z [(2)] ($\times 10^6$ m)	z_{max} [(3)] ($\times 10^6$ m)	Entered cavity?	Reached cavity bottom?
(0,0,0)	4	-8.82	0.01	Yes	No
(0,0,-H_R)	4	-9.75	-9.66	No	No
(W_R/2, W_R/2, -H_R)	4	-9.90	-9.67	No	No

[(2)] D = diffusion coefficient of particle (10^{-10} m^2/s); [(3)] N_1 = Time of entry into cavity; [(4)] N_2 = Time to reach cavity bottom
[(1)] N_3 = Total steps; [(2)] z (see Figure 1a) = distance ($\times 10^6$ m) after N_3 steps ; [(3)] z_{max} = Maximum z reached during N_3 steps.

Condition 2. Referring to Table 1 above tRNA is clearly separated by dimension from AARS on one side and from AMP, ATP, and any of the AAs on the other. Thus filtration can be done with nanopores of sizes given in Table 1.
Condition 3. In the living cell deacylation is done by an enzyme to correct charging errors [24]. 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 [28]. Here non-enzymatic deacylation with NaOH is assumed; see [27] for its use in the deacylation of cysteine-tRNA.
Condition 4. Consider Equation (7). 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. With a signal-to-noise ratio SNR = V_{AN} / V_{RMS} a sufficiently high SNR can ensure that the blockade current due to G can always be distinguished from I_{RMS}. In [34] 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 Figure 7 below and the Supplement for more information.

Processing in parallel with N or more AA molecules and N different tRNAs
The sequence of steps given above can be done in parallel with N different tRNAs and N copies of AA, filters, and e-cells, as shown in Figure 3. With N = 20 superspecificity of tRNA guarantees that AA will be identified by one of the 20 units.

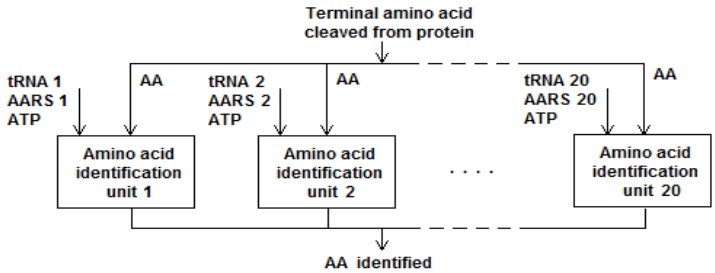

Fig. 3 Parallel architecture for single AA identification

Processing in series if only a few molecules are available
If only a few molecules of AA are available (as happens if the sample comes from single cells) the processing can be done in series, as shown in Figure 4.
.

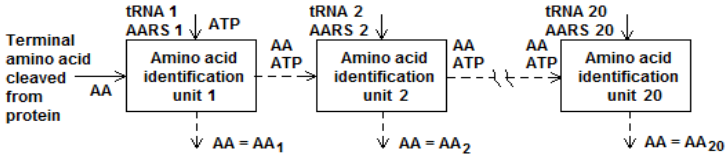

Fig. 4 Series architecture for single AA identification

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, let 1 of M_1 tRNA molecules get charged in the cavity and 1 of M_2 charged tRNA molecules are deacylated. In a parallel implementation with 20 units, a total of 20 M_1M_2 AA molecules in the sample and 20 M_1M_2 tRNA molecules can identify AA with probability 1. With M_1=M_2=10 (corresponding to charging and deacylation rates of 10%), this is about 3 zeptomoles, which is comparable to the molar volumes of samples used in a recent label-based method [10].

Protein sequencing
The full peptide or protein sequence can be obtained by applying the above procedure to terminal residues cleaved in succession from a peptide/protein. Cleaving may be done with ED, which is known to be a harsh process and is

not suited to the present approach because ED produces derivatized AAs. Less harsh alternatives that produce free AAs, such as carboxyl peptidases [35] and aminopeptidases [36] are more appropriate; see [17] for a brief discussion. In [37] an enzyme designed on a computer and named Edmanase that cleaves residues at the N-end of a protein is described.

3 A two-step procedure to identify AA

AA can also be identified from the AMP molecule that is released during the charging process. The latter is allowed to translocate through a nanopore and the blockade due to it measured. No filtering is required so the number of steps is reduced to two. The tradeoff is that the AMP blockade must be distinguished from blockades due to some of the other reactants (ATP, free AAs) translocating through the pore. While this introduces some error, the error rate is still < 10%; this is discussed below. Figure 5 shows the setup. The cavity is connected through a pore of appropriate size to an e-cell so that after charging has occurred the smaller molecules (AMP, ATP, AA) can exit the cavity into the *cis* chamber of the e-cell. Entry is aided by hydraulic pressure on the reservoir, with the *trans* chamber open to atmospheric pressure. This allows any of the three analyte types to cause measurable blockades. The pore size of 3 nm (Table 1) ensures that AARS and tRNA (charged or not) cannot enter the e-cell. This design considerably simplifies both the hardware and the process of recognizing current blockades due to AMP, ATP, or AA.

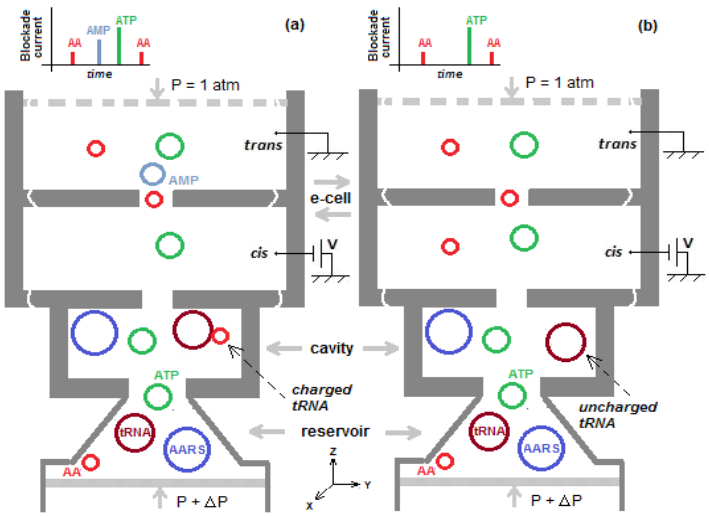

Fig. 5. Integrated device for amino acid identification. Reservoir-cavity structure at bottom is similar to Figure 2. Electrolytic cell (e-cell) above it is connected through pore; *trans* chamber in e-cell is open to atmospheric pressure. Hydraulic pressure applied to reservoir allows ATP, AMP, and AA to pass into *cis* chamber and thence translocate to *trans*, thereby causing blockade. Occurrence of AMP blockade signals presence of AA, whose identity is known from identity of tRNA.

Step 1: Charging of tRNA

This is similar to Step 1 in the four-step procedure with one small difference: the hydraulic pressure applied at the reservoir end is > 1 atm to enable the small molecules to exit the cavity, enter *cis*, and translocate to *trans*.

Step 2: Identification of AA from AMP blockade

After AA, AMP, and ATP enter the *cis* chamber of the e-cell connected to the cavity (Figure 1b and 1c) they translocate into *trans* under the influence of hydraulic pressure and cause current blockades. If tRNA is charged the translocating analytes are AMP, ATP, and free AAs; if there are no charged tRNAs only free AAs and ATP pass through.

Conditions for correct identification of an amino acid

The following conditions have to be satisfied in order for the above procedure 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) An AMP blockade can be distinguished from blockades due to ATP or free AAs that enter the e-cell below the cavity.

The first condition is the same as in the four-step procedure. However the identification procedure is different as it is based on measuring AMP blockades. Unlike in the earlier procedure, which is based on observation of a blockade due to AA without any other analytes being involved, here AMP blockades may be accompanied by blockades due to free AA and ATP molecules. Hence the second condition is better understood in terms of the errors that may occur when an AMP blockade is measured. This is addressed next.

Error analysis

The identification process can be analyzed for potential *false negatives* and *false positives*. A false negative occurs if the tRNA is not charged and ATP were to cause an AA-sized blockade, or the ATP looks like an AMP and the AMP looks like an ATP. It is sufficient to consider the former possibility as it has the higher probability. A false positive occurs if the tRNA is not charged and AA were to cause an AMP-sized blockade. Figures 6a and 6b show the variation of the corresponding two probabilities P_{FN} and P_{FP} for each of the 20 amino acids for four values of σ (0.05μ, 0.1μ, 0.15μ, 0.2μ), where μ and σ are the mean and standard deviation of a normal distribution of analyte volume. The worst case occurs with W (Tryptophan), the largest volume AA, with a 17% false negative probability when σ = 0.2μ. If the two types of errors are averaged over all 20 amino acids weighted by their frequency of occurrence in the human proteome (Uniprot Id UP000005640; 11326153 residues) the weighted error probability is 9.5% for σ = 0.2μ; see Figure 6c.

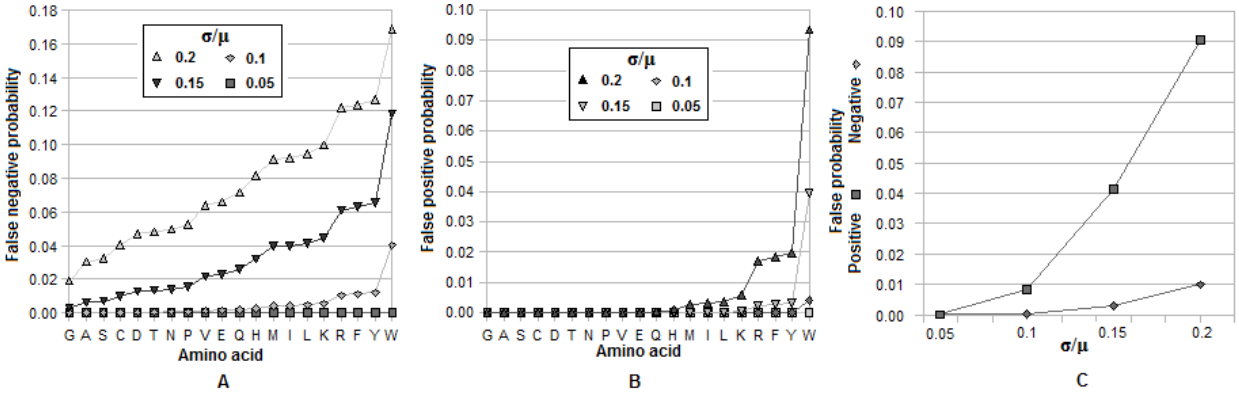

Fig. 6 Error probabilities for the standard 20 amino acids. (a) False negatives; (b) False positives; (c) Both over the full set of proteins in the human proteome weighted by frequency of occurrence of each amino acid

Figure 7 shows the probability that a measured blockade for the smallest AA, Glycine, is less than the noise threshold $Th_{noise} = 2\ I_{RMS}$ for different values of $\sigma_{AA=G}$, where σ_{AA} is the standard deviation in the volume of AA.

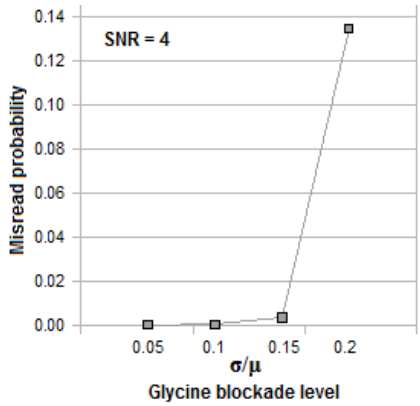

Fig 7. Probability of 2*baseline RMS noise (I*RMS*) in nanopore current being mistaken for measured blockade due to smallest volume AA (G) with a signal-to-noise ratio (SNR) of 4 for different values of standard deviation σ relative to mean volume μ of G.

4. Discussion and Conclusions

The amino acid identification procedures described above have several advantages: 1) By shifting most of the burden of AA identification to tRNAs and their cognate AARSs instrumentation precision and noise no longer have the critical importance they do in methods based on electrical, optical, or mass measurements; 2) The absence of labeling makes the method much simpler; and 3) Such properties as pore length, electro-osmotic effects, and hydrophobicity [16] are not important. The only major consideration is that in the e-cells used the pore diameter should be sufficiently small that a tRNA (charged or not) or AARS cannot enter the pore.

Some general comments and notes on possible improvements to the identification procedure are given next.
1) In the 2-step procedure the size of the AA blockade can be correlated with the identity of the tRNA that identifies AA.
2) In the 4-step procedure the number of filtrations can be reduced from 2 to 1 if deacylation does not have an adverse effect on AARS. Thus there is no need to separate AARS from tRNA.
3) Note that re-use of tRNA and AARS is almost total; ATP also can be recovered by filtration. Also once the tRNA and AARS for a processing unit are loaded there is no need to reload so the unit can be used an indefinite number of times.
4) AA blockade level analysis can potentially be used to get post-translational modification (PTM) information.
5) The series architecture in Figure 4 is reminiscent of chromatography columns.
6) Since each tRNA/AARS pair is isolated from any other every unit, each unit can be optimized for the particular AA: thus hydraulic force and electrical potential can be pre-selected by magnitude and sign to match the target AA's volume, electrical charge, and polarity (somewhat like looking at the answer in the back of the book).
7) In this approach a physical measurement is done only in the last stage with a nanopore, even then only nominal precision is required. All preceding stages involve chemical binding (tRNA charging; deacylation) or physical separation (filtering); no quantitative analysis is done or required.

The following points are worth noting:
1) Superspecificity is for all practical purposes error-free. As a result the extreme demands that are a characteristic of nanopore-based sensing are diluted to such an extent that the role of the nanopore becomes secondary.
2) In the first procedure it suffices to know if a blockade has occurred in the e-cell. In the second it is sufficient to know if an AMP blockade has occurred. In both cases there is no attempt (or need) to distinguish between any of the 20 proteinogenic AAs, which is rather ironic.
3) No information about the target protein is used or needed anywhere. This makes both the procedures presented truly *de novo* in character. Currently the only practical *de novo* protein sequencing method is MS, but that is the result of applying machine learning techniques to the spectral data: the *de novo* capability is a computational feature, not an analytical one. At present there appear to be only four other methods with *de novo* characteristics: [11], [17] and [18], which are theoretical methods like the work presented here, and [19], which is a practical method but not a fully proven one to date.

Supplement

File with simulation details, supporting data, notes on error analysis, discussion, and additional references.

References

[1] R. J. Simpson. Proteins and Proteomics: A Laboratory Manual, CSHL Press, 2008.
[2] T. Rabilloud and C. Lelong, “Two-dimensional gel electrophoresis in proteomics: a tutorial.” J. Proteomics 74, 1829-1841, 2011.
[3] E. de Hoffmann and V. Stroobant. Mass Spectrometry: Principles and Applications, 3rd edn., Wiley, 2007.
[4] S. Xie1, C. Moya, B. Bilgin, A. Jayaraman, and S. P. Walton, "Emerging affinity-based techniques in proteomics", Expert Rev Proteomics, 6, 573–583, 2009. doi:10.1586/epr.09.74
[5] Yusko EC, Bruhn BR, Eggenberger O, Houghtaling J, Rollings RC, Walsh NC, Nandivada S, Pindrus M, Hall AR, Sept D, Li J, Kalonia DS, and Mayer M. “Real-time shape approximation and fingerprinting of single proteins using a nanopore”. Nature Nanotech., 2016. doi 10.1038/nnano.2016.267
[6] Bell NAW and Keyser UF “Specific protein detection using designed DNA carriers and nanopores”, J. Am. Chem. Soc. 2015. doi 10.1021/ja512521w
[7] M. Mann and N. L. Kelleher, "Precision proteomics: The case for high resolution and high mass accuracy", PNAS, 105, 18137, 2008.
[8] G. Sampath, "Protein fingerprinting with digital sequences of linear protein subsequence volumes: a computational study", J. Biosci. 44-54, 2019. doi: 10.1007/s12038-019-9863-9
[9] 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.

[10] J. Swaminathan, A. A. Boulgakov, E. T. Hernandez, A. M. Bardo, J. L. Bachman, J. Marotta, A. M. Johnson, E. V. Anslyn, and E. M. Marcotte. “Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures”. Nature Biotechnol. 36, 1076-1082, 2018.
[11] S. Rodriques, A. Marblestone, and E. Boyden. “A theoretical analysis of single molecule protein sequencing via weak binding spectra”. PloS One 14, e0212868, 2019. doi: 10.1371/journal.pone.0212868
[12] Borgo B, Havranek JJ. Motif-directed redesign of enzyme specificity. Protein Science., 23, 312–320, 2014. doi: 10.1002/pro.2417
[13] J. E. Reiner, A. Balijepalli, J. W. F. Robertson, J. Campbell, J. Suehle, and J. J. Kasianowicz. “Disease detection and management via single nanopore-based sensors”. Chem. Rev. 112, 6431-6451, 2012.
[14] M. Chinappi and F. Cecconi, “Protein sequencing via nanopore based devices: A nanofluidics perspective”, J. Phys. Cond. Matter **30**, 204002, 2018.
[15] L. Restrepo-Pérez, C. Joo, and C. Dekker. “Paving the way to single-molecule protein sequencing”. Nature Nanotechnology 13, 786-796, 2018. https://doi.org/10.1038/s41565-018-0236-6
[16] W. Timp and G. Timp, "Beyond mass spectrometry, the next step in proteomics", Sci. Adv. 6, eaax8978, 2020.
[17] G. Sampath. “Amino acid discrimination in a nanopore and the feasibility of sequencing peptides with a tandem cell and exopeptidase”. RSC Adv. 5, 30694–30700, 2015.
[18] P. Boynton and M. Di Ventra. “Sequencing proteins with transverse ionic transport in nanochannels”. Sci. Rep. 6, 25232, 2016.
[19] Y. Zhao, B. Ashcroft, P. Zhang, H. Liu, S. Sen, W. Song, J. Im, B. Gyarfas, S. Manna, S. Biswas, C. Borges, and S. Lindsay. “Single-molecule spectroscopy of amino acids and peptides by recognition tunneling”. Nature Nanotechnol. 9, 466–473, 2014.
[20] E. Kennedy, Z. Dong, C. Tennant, and G. Timp. “Reading the primary structure of a protein with 0.07 nm3 resolution using a subnanometre-diameter pore”. Nature Nanotechnol. 11, 968–976, 2016.
[21] 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., 38, 176–181, 2020. doi: 10.1038/s41587-019-0345-2
[22] 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
[23] M. Ibba and D. Söll. “The renaissance of aminoacyl-tRNA synthesis”. EMBO Reports, 2, 382-387, 2001.
[24] G. Bergtrom, Cell and Molecular Biology, 3rd edn. University of Wisconsin, Milwaukee, 2018.
[25] D. Goodsell. “Aminoacyl-tRNA Synthetases”. PDB Molecule of the Month, April 2001. doi:10.2210/rcsb_pdb/mom_2001_4
[26] 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. 13, 3048–3052, 2013.
[27] Jakubowski H “Editing function of Escherichia coli cysteinyl-tRNA synthetase: cyclization of cysteine to cysteine thiolactone”. Nucl. Acids Res. 22, 1155-1160, 1994.
[28] Gomez MAR, Ibba M. “Aminoacyl-tRNA synthetases”. RNA (online), April 17, 2020.
[29] Takita T. “The structure and functions of the lysyl-tRNA synthetase of bacillus stearothermophilus”. Dissertation, Kyoto University, March 1996. doi: 10.11501/3110616
[30] Brady KT, Reiner JE. “Improving the prospects of cleavage-based nanopore sequencing engines”. J. Chem. Phys. 143, 074904, 2015.
[31] Sampath G “DNA sequencing with stacked nanopores and exonuclease: a simulation-based analysis”. Electrophoresis 37, 2429-2434, 2016. doi: 10.1002/elps.201600049
[32] Pedone D, Langecker M, Abstreiter G, Rant U “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. 11:1561–1567, 2011.
[33] Liu X, Skanata M, Stein D (2015) Entropic cages for trapping DNA near a nanopore. Nature Communs. 6, 6222, 2015.
[34] A. Fragasso, S. Schmid, and C. Dekker. “Comparing current noise in biological and solid-state nanopores”. bioRxiv.org, 2019. doi: 10.1101/866384
[35] K. Breddam and M. Ottesen, “Determination of c-terminal sequences by digestion with serine carboxypeptidases: the influence of enzyme specificity”. Carlsberg Res. Commun. 52, 55-63, 1987.
[36] A. Taylor, “Aminopeptidases: structure and function”. FASEB J. 7, 290-298, 1993.
[37] B. Borgo and J. J. Havranek. "Computer-aided design of a catalyst for Edman degradation utilizing substrate-assisted catalysis". Protein Sci. 24, 571–579, 2015.