On sequencing a protein with nanopores, tRNAs, and RNAs with matching terminal codons

G. Sampath

P. O. Box 7849, J P Nagar, Bengaluru 560078, India

Abstract. A method for sequencing a protein from a codon sequence is proposed. An unfolded protein molecule is threaded through a nano-sized pore in an electrolytic cell carboxyl end first and held with a voltage such that only the first residue is exposed in the *trans* chamber of the cell. A tRNA molecule in *trans* with matching anticodon for the residue binds itself to the latter in the presence of suitable catalysts. It is then cleaved and transferred to an extended electrolytic cell with N pores, $40 \leq N \leq 61$, in N individual *cis* chambers and a single *trans* chamber. Each pore holds an RNA molecule ending in a unique codon that is held exposed in the *trans* chamber. In the presence of suitable catalysts the anticodon in the transferred tRNA binds with the codon of a matching RNA molecule. By reversing the voltages in the extended cell every RNA molecule except the one to which the transferred tRNA is bound retracts into its *cis* chamber, this identifies the residue unambiguously. The detected residue in the first cell is cleaved and the process repeated. Unlike in other nanopore-based methods, it suffices to detect the occurrence of a current blockade without having to measure the pore current precisely. A simplified but more time-consuming version that uses only the first cell is also described. In either case no *a priori* information about the protein is needed so *de novo* sequencing is possible. A feasibility analysis of the proposed scheme is presented.

Keywords: protein sequencing; nanopore; tRNA; RNA; codon; amino acid charging; optical tag

1. Introduction

Biopolymer sequencing occupies a central place in the study of biological organisms, it may be focused on research or the diagnostic and therapeutic value of the information contained in the polymer sequence. DNA sequencing is now a mature area [1] and is accomplished by a number of technologies, starting with Sanger sequencing and proceeding to next generation techniques that make use of highly automated processes [2]. On the other hand protein sequencing is not as advanced, largely because there are 20 amino acids to work with, compared to four bases for DNA. It is currently done with one of the following methods: Edman degradation (ED), gel electrophoresis (GE), or mass spectrometry (MS) [3-5].

More recently nanopores have been used to sequence DNA [6]. Such methods are beginning to show promise and are set to compete with the more established methods mentioned above [7]. In contrast, nanopore-based protein sequencing is still in its infancy; a variety of methods, some in practice [8-11], others in theory [12-15], are known.

1.1 *The present work*

This report introduces a hypothetical sequencing scheme in which the sequence of amino acids in a protein is obtained indirectly from a sequence of codons that code for the amino acids. An electrolytic cell with a nanopore is used to hold the unfolded linear protein sequence in place while a tRNA is attached to the first residue of the polymer. This is followed by transfer of the tRNA to a second modified electrolytic cell with multiple nanopores each holding an RNA molecule ending in an amino-acid-coding codon. The residue is identified when the anticodon in the tRNA binds with the matching terminal codon of one of the RNA molecules. The identified residue in the first cell is cleaved, the polymer advanced by one residue through the pore, and the process repeated until the protein has been sequenced. The proposed scheme is described in detail and its practical feasibility discussed.

In this proposed approach, which combines the use of nanopores with conventional reagent-based methods, a major problem encountered by many existing methods for protein sequencing, namely the inability to control peptide location and orientation, is considerably mitigated because the protein is held stationary in a nanopore while a single terminal residue is exposed to enzymes and other reagents. Importantly no (precise) measurements of the nanopore current are required, it is sufficient to just detect the occurrence of a current blockade. If the catalysts involved work as required (or can be modified for the purpose), the proposed method could lead to error-free *de novo* protein sequencing.

2. Nanopores, tRNAs, and RNA molecules with terminal codons

An electrolytic cell [16] consists of two chambers, named *cis* and *trans*, which contain an electrolyte, typically NaCl or KCl, and are bridged by a thin membrane containing a nanopore. An electrical potential V applied between *cis* and *trans* ionizes the electrolyte and results in current flow through the pore. With KCl as the electrolyte, K^+ ions flow toward the cathode and Cl^- toward the anode. If a charged analyte molecule, such as DNA, RNA, or protein, is introduced into *cis*, it translocates through the pore in an appropriate direction. Both DNA and RNA carry a uniform negative charge along their backbone, so they flow from *cis* to *trans* in Fig 1 below. In contrast, seven of the standard 20 amino acids, namely D, E, K, R, H, C, and Y, carry a negative or positive charge whose value depends on the pH of the electrolyte [17]. This means that the net charge carried by a protein depends on the sequence of residues and also determines the direction of flow of the protein. Crucially, translocation of DNA, RNA, or protein through the pore causes a reduction in the pore current that varies with the monomer passing through the pore. This monomer-specific level of current blockade serves as the basis for current approaches to nanopore-based sequencing. Thus thinner pores offer better discrimination. In contrast, in the scheme described here the blockade level is not important, just the occurrence of a blockade is sufficient (no current measurements are made). Pore thickness is also not a consideration. In fact thicker pores may be preferable as they provide better translocation control (see Section 5 below).

Two kinds of electrolytic cells are considered: a) a standard electrolytic cell (described in the previous paragraph) to hold the unfolded protein in place (Fig. 1(a)); and b) an electrolytic multicell with a single *trans* chamber and N *cis* chambers each of which holds in place an RNA molecule ending in a unique codon (Fig. 1(b)). Here N is the number of unique tRNAs. It is typically 40 but may be as high as 61 (the maximum number of coding triplets in the genetic code) [18].

The first cell is used to advance the protein molecule one residue at a time through the pore and find a tRNA that codes for that residue. The coding tRNA is then transferred to the second cell where it is identified unambiguously.
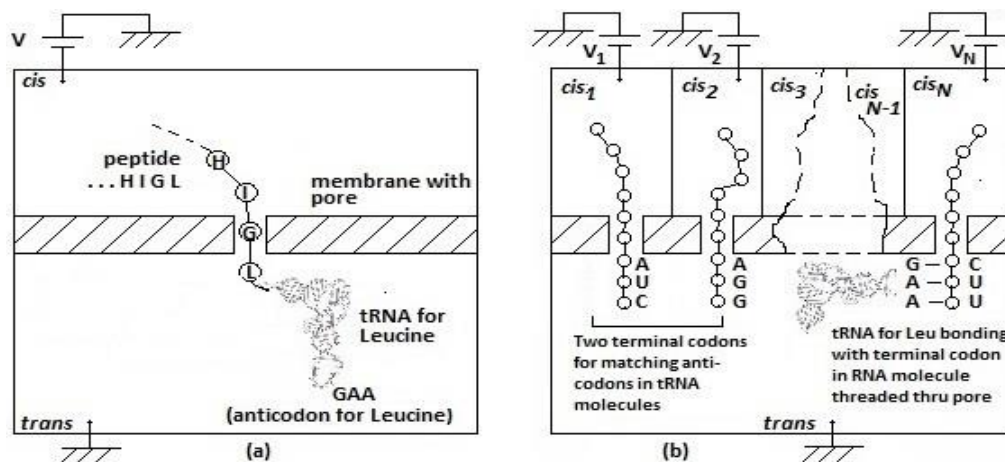

Fig 1. Electrolytic cells for sequencing a protein molecule (not to scale)

3. Rationale for the proposed protein sequencing procedure

Consider the following requirements for a protein sequencing procedure based on the elements described in Section 2.

1) The terminal residue of the protein molecule at the carboxyl end must be exposed so that it can be bound to its matching tRNA [18]. Such exposure can be done by threading the unfolded protein molecule from the *cis* chamber of an electrolytic cell through the nanopore in the cell's membrane so that the first residue is in the *trans* chamber and all its successors are in the pore and the *cis* chamber. In Fig. 1(a) the first residue is L (Leucine); it is held in *trans* just outside the pore. Such controlled exposure of the first residue can be achieved by adjusting the voltage V across *cis* and *trans*.

2) If all tRNA types and other catalyzing agents for amino acid charging are present in *trans* then with high probability the exposed residue will be charged, that is, a tRNA with an anticodon matching the codon for the

amino acid will attach itself at its 3' end to the residue. In Fig. 1(a) this tRNA has anticodon GAA; the corresponding codon CUU codes for the exposed residue Leucine (L).

2) Once the tRNA for the exposed residue has attached itself the remaining (unattached) tRNAs in *trans* need to be flushed out so that the coding tRNA can be cleaved with an appropriate enzyme and transferred to the second electrolytic cell (multicell) for identification.

3) The coding tRNA from the first cell can be identified if it is made to attach itself at its anticodon end to a matching codon and the latter identified in some way. For this all possible codons must be made available to the tRNA. Consider how this can be done. The electrolytic multicell has N independent *cis* chambers each with a pore through a single common membrane that separates them from a single common *trans* chamber. N RNA molecules each with a unique codon at its end can be threaded from a *cis* chamber through the corresponding pore such that the unique triplet is exposed in *trans* while all the succeeding bases are in the pore and the *cis* chamber. Such selective exposure can be achieved by controlling the voltage across each *cis-trans* pair in the electrolytic multicell. In Fig. 1 (b), the matching codon for the tRNA (with anticodon GAA) that is transferred from the first electrolytic cell is CUU.

4) If the tRNA is transferred from the first cell to the *trans* chamber of the multicell, it is exposed to all possible N codons there. Then in the presence of appropriate catalysts the tRNA will with high probability attach itself to the matching terminal codon of one of the RNA molecules. (There could be more than one such matching RNA molecule because of the redundancy inherent in the genetic code [18], but there is no ambiguity in the identification because the tRNA will bind to only one of the matching RNA molecules.)

5) To identify the terminal codon to which the tRNA molecule attaches itself the voltages V_1, V_2, ..., V_N are reversed so that the RNA molecules retract into their *cis* chambers. Every RNA molecule except the one to which the tRNA has attached itself will translocate fully into its *cis* chamber and the corresponding pore current will rise to the free pore current level. The one with the tRNA cannot do so because the tRNA will block the pore and the current will remain at the blockaded level. (The blockade level may even increase.) It suffices to know whether a pore current is at the blockaded level or not, there is no need to know its exact value. Identification of the codon for the tRNA, and hence of the residue, is unambiguous.

6) The first residue in the protein molecule in the first electrolytic cell is cleaved with an appropriate peptidase (or other chemical). (The cleaving agent to use is known exactly because the residue has been identified.) Following this the next residue in the protein sequence is advanced into the *trans* chamber of the first cell, and the process repeated.

Note that this is essentially a *de novo* approach as no *a priori* knowledge of the protein in any form is used.

The protein sequencing procedure is given next.

4. Protein sequencing procedure
The following sequence of steps is repeated for every residue in the target protein.

Identification procedure for a single residue

Step 1.   Advance protein molecule from carboxyl end through pore in first cell (Fig 1(a)) with voltage V set such that first residue is outside pore in *trans* chamber of cell while rest of polymer remains in pore and *cis* chamber.

Step 2.   Add all possible tRNA molecules to *trans* chamber of first cell and catalysts for attaching 3' binding site (CCA terminus) of tRNA to corresponding amino acid. tRNA for terminal residue (Leucine in example above) binds, all other tRNAs remain in *trans* solution.

Step 3.   Flush these remaining tRNAs out of *trans* while refilling with electrolyte.

Step 4.   Prepare second cell (Fig. 1(b)). Thread RNA molecule ending in unique triplet corresponding to codon for amino acid i through pore i in electrolytic multicell. Set voltages V_i such that terminal triplet in RNA molecule is in *trans* chamber of electrolytic multicell and trailing bases are in pore i and cis_i. In each case pore current is lower than free pore current because of blockading effect of RNA molecule in pore.

Step 5.   Cleave tRNA from protein terminal residue with catalysts. Flush *trans* solution containing cleaved tRNA (and catalysts) out to *trans* chamber of electrolytic multicell while refilling *trans* of electrolytic cell with electrolyte.

Step 6.   Add catalysts to *trans* of electrolytic multicell such that anticodon of transferred tRNA bonds with terminal codon (triplet) of j-th RNA molecule for some j.

Step 7.   Flush out *trans* of electrolytic multicell while refilling with electrolyte. This removes all catalysts that were introduced in Steps 5 and 6.

Step 8.   Set all V_is to values that will cause all threaded RNA molecules to retract into their *cis* chambers. This results in end of current blockade due to RNA molecule in every pore except the one to which tRNA has attached itself (j in Step 6). In latter case current blockade continues.

Step 9.   Read out pore number (j) for pore with continued blockade. This identifies terminal residue in protein.

Step 10. Flush out *trans* chamber of electrolytic multicell and refill with electrolyte. Re-thread RNA molecules through every pore in electrolytic multicell.

Step 11. Add peptidase/chemical to *trans* of electrolytic cell to cleave first residue from protein. (Cleaving agent is specific because terminal residue has been identified in Step 9.)

Alternatively Step 11 can be combined with Step 5 to eliminate cleaving of the tRNA. The exposed residue can be cleaved in Step 5 so that the charged tRNA separates and is transferred to the multicell by flushing the *trans* chamber of the first cell into the *trans* chamber of the multicell. (In this case since the tRNA has not yet been identified the required cleaving agent is not known specifically so more than one cleaving agent may have to be tried.)

Flushing out a *trans* chamber minimizes the possibility of spurious reads. For example, a residual enzyme from Step 6 could block one of the pores other than the one targeted by the tRNA, resulting in a misread. Similarly Step 3 ensures that only the coding tRNA gets transferred to the *trans* chamber of the electrolytic multicell in Step 5. Notably, all flushed out reagents can be reused in the next cycle, which also makes for a greener approach.

5. Feasibility analysis

In order for the proposed scheme to be feasible, several conditions must be met. They can be divided into three categories: 1) use of appopriate catalysts; 2) event detection; and 3) polymer positioning control.

1. *Catalysts required*

Four major chemical reactions are involved in the procedure given in Section 4:

1) attaching a tRNA molecule at its 3' end to the carboxyl end of the exposed residue in the first cell;

2) cleaving the above tRNA molecule from the residue;

3) binding a tRNA to the exposed terminal codon of an RNA molecule in the second cell;

4) cleaving the exposed residue from the protein molecule in the first cell.

The first three reactions have roughly corresponding ones that occur in a biological cell during translation. These latter reactions are extremely complicated and involve a large number of complex interactions involving a host of enzymes and other factors with codon-specific tRNAs to facilitate translation of an mRNA into a peptide sequence [19]. Thus attaching a tRNA to the terminal residue of the growing peptide involves the correct aminoacyl tRNA synthetase enzyme and ATP in a two stage reaction. Cleaving a tRNA from the last residue that was attached to the peptide is part of a more complicated reaction involving the enzyme peptidyl transferase and several other factors. This requires attaching an incoming residue on a tRNA molecule to the previously attached residue in the growing peptide while detaching the previously attached tRNA from the latter. Binding of the anticodon in a tRNA to the matching codon of the mRNA is based on Chargaff base-pair matching but is rendered more complicated by the tight environment of the ribosome in which a host of ribosomal proteins and RNAs are involved [19].

In contrast, the proposed scheme is considerably different and probably easier to work with because the environment is a much simpler and artificial one under the control of the designer. Thus whereas a biological cell is dynamically regulated by a number of concomitant processes occurring in a restricted and crowded space, in the present case the positions of the exposed residue in the first cell and of the terminal codons in the N RNA molecules in the second cell are all fixed and can be targeted in a controlled manner. The chemicals required for a reaction can be restricted to the ones required for that reaction. For example, one consideration is that the exposed residue be charged correctly: the right tRNA must be bound at its 3' end to the residue at the carboxyl end. Here one can be guided by how this is done in a biological cell, where, to minimize charging errors, tRNAs

and the appropriate aminoacyl tRNA synthetases are present in the correct proportion in the cell. (The accuracy of amino acid charging in cell metabolism is so high that aminoacyl tRNA synthetases are said to be 'superspecific'.) Another consideration is that while the exposed residue is free at the carboxyl end, the amino end is not as it is bonded to the next residue in the protein (which is inside the pore). If it turns out that this has an adverse effect on charging other ways may be explored. One possibility is cleaving the exposed residue before adding tRNAs and catalysts. This would require transferring all the tRNAs (including the one attached to the free amino acid) to the multicell, making the identification process there more complicated but not inordinately so. As described below, if the amino acid carries a distinctive optical tag its location in the multicell, and hence its identity, can be known precisely. Another possibility is to modify the tRNA and/or the corresponding synthetase to ensure that charging occurs even while the exposed residue remains bonded to its successor in the protein molecule in the pore.

Alternatively tRNA and synthetase pairs can be tried out in series until charging occurs (this event can be detected optically as described below). Writing the standard set of amino acids as **AA** = [G, A, S, C, D, T, N, P, V, E, Q, H, M, I, L, K, R, F, Y, W], a maximum of 20 cycles, one per amino acid, would be required. The residue may be identified in the first cycle (as G, Glycine) in the best case and in the 20th cycle (as W, Tryptophan) in the worst case, with an average of 10 cycles (for E, Glutamic Acid). In this case only the first electrolytic cell is needed. This is a much simpler procedure but more time-consuming. However, it is fairly easy to automate and requires only one color for optical detection (only the tRNA needs to be tagged, see below).

The modified procedure for a single residue is given next, there are only 4 steps. It is repeated until the protein has been sequenced.

Modified identification procedure for a single residue
Step 1.   Advance protein molecule from carboxyl end through pore in first cell (Fig 1(a)) with voltage V set such that first residue is outside pore in *trans* chamber of cell while rest of polymer remains in pore and *cis* chamber.
Step 2'.   For every amino acid type i, i = 1 to 20, repeat the following until residue is identified:
    Add tRNA molecule i corresponding to amino acid type i, as well as other required catalysts, to *trans* chamber of first cell for attaching 3' binding site (CCA terminus) of that tRNA to exposed residue. After charging has occurred flush remaining free reactants out of *trans*.
    (This loop is exited for some i in the range 1 to 20, the residue is identified as amino acid type i. In the example above a tRNA molecule for Leucine binds in the 15-th pass.)
Step 3'.   Add peptidase/chemical to *trans* to cleave first residue (with its matching tRNA attached to it) from protein. Cleaving agent is specific because terminal residue has been identified in Step 2'.
Step 4':   Flush cleaved residue with attached tRNA and all other reagents out of *trans* while refilling with electrolyte.

The fourth reaction does not have a counterpart in the mRNA-to-peptide translation process in a biological cell and is probably the easiest one of the four to implement. Proteolysis is a widely studied process and there is a vast literature on the subject; comprehensive reference works are available [20,21]. In the present context the objective is limited to cleaving the first exposed residue of a protein molecule in the pore of the first electrolytic cell. Some applicable methods are discussed in [12] in the context of nanopore-based protein sequencing using a tandem electrolytic cell. More recently an enzyme designed on a computer and named Edmanase that cleaves residues at the N-end of a protein has been described [22]. In [23] naturally occurring proteins that recognize specific N-terminal amino acids have been studied for possible use in peptide sequencing, leading to variants of the protein ClpS that can recognize N-terminal F (Phenylalanine) and W (Tryptophan).

2. Event detection

The procedure in Section 4 and the modified one given above involve up to four distinct events that occur in sequence:
1) attachment of tRNA to the exposed residue in the first electrolytic cell;
2) detachment of the same tRNA from the residue;
3) binding of the transferred tRNA in the second electrolytic multicell to the terminal codon of some RNA molecule; and

4) cleaving of the exposed residue in the first cell.

The times of their occurrence are instrumental in driving the next step. This does not present a problem here because the sequence of steps is under the control of the designer. Also actions in the two cells can be pipelined for faster sequencing.

All four events can be detected optically by attaching optical tags [13,24,25] to the relevant molecules. Such detection is facilitated by the fact that the physical positions associated with them are fixed and stationary (or can be related in a simple way to other fixed positions). Therefore the position of a tag following an event is easily determined. For example when the exposed residue is still attached to the protein molecule the location of the optical signal of its tag is fixed. However after it is cleaved, the tag location is no longer stationary and this is sufficient to signal the completion of cleaving. Similarly when the anticodon in a tagged tRNA has been bound to the matching codon of an RNA molecule in the second cell, the optical signal location is more or less determinate. If the RNA molecules are also tagged then the pair of locations of the targeted RNA tag and the tRNA tag can be used to identify the codon. In this case there is no need to retract the RNA molecules in the second cell into their *cis* chambers (Step 8 in Section 4). At most three colors are needed for optical detection (compare with [13] and [24]).

3. *Polymer translocation control*

The analyte molecule (DNA or protein) translocates rapidly through the pore of an electrolytic cell thus making it difficult to measure the change in the current blockade level due to individual monomers [26]. Currently available detector bandwidths do not support such measurements. In the present case, however, the only consideration is to hold the protein in the pore in the electrolytic cell and the RNA molecules in the electrolytic multicell such that only the terminal residue and the terminal codons are exposed in the respective *trans* chamber while the molecule (protein in electrolytic cell or RNA in electrolytic multicell) remains stationary.

The voltage levels required to hold a polymer stationary can be better understood from the physics of polymer translocation. There are two forces at work: diffusion and electrophoresis [16]. (Electro-osmotic forces due to charges on the pore interior surface are neglected in this analysis as their relevance is minimal.) Diffusion is always from source chamber to destination through the pore. It is always present, even when there is no applied voltage. The required holding voltage is the value that negates this diffusive force, its sign depends on the charge carried by the polymer. Whereas DNA and RNA carry a uniform negative charge along the backbone, most proteins carry a weak net charge that may be positive or negative. In the latter case better control is possible by attaching a trailing homopolymer polyZ where Z is a charged amino acid with appropriate polarity.

Other approaches to polymer translocation control include use of the following: (i) a room-temperature ionic liquid (RTIL), which is a high viscosity electrolyte that can slow down an analyte by a factor of ~200 [27]; (ii) an opposing hydraulic pressure field [28]; (iii) an enzyme ('unfoldase') to unfold the protein molecule before it enters the pore and then slow its passage [29]; (iv) ligands attached to the protein or the pore [30]. The first two can be used for both RNA and protein.

Most of the voltage drop in an electrolytic cell is across the pore [16]. The drop in the *cis* and *trans* chambers is only around 2% so the motive force in these chambers is mostly diffusion. Because of this an analyte molecule needs to be drawn into the pore by some agency such as dielectrophoretic trapping [31]. With proteins a possible solution is to use a sodium dodecyl sulphate (SDS) sheath to give the protein a uniform negative charge that gets stripped out as the protein is drawn into the pore [9].

The protein molecule must enter the pore in the electrolytic cell carboxyl end first as the tRNA attaches to this end. To ensure entry at the correct end the amino end can be capped with a small molecule of a size large enough that it cannot enter the pore. This also requires a polyZ trailer of length b, where Z is a selected amino acid and b*(length of Z) > length of pore. Similarly an RNA molecule in the electrolytic multicell has to enter a pore at the 3' end. Once again capping can be used for this. An alternative approach in both cases would be to reverse the roles of *cis* and *trans*, which can effectively be achieved by reversing the voltage polarity. Entry into a pore at the wrong end can be detected (optically) when the desired event (charging of tRNA, binding of tRNA to codon) does not occur within some expected time.

Tethering offers a workable and practical solution to the translocation control problem with both protein and RNA. Reference 10 gives an example of its use in the identification of residues in peptides via total internal reflection fluorescence (TIRF) microscopy. Consider the following:

1) Let the protein sequence $R_1 R_2 ... R_n$ going from N-terminal to C-terminal be extended by a poly-X trailer of

length m, where X is one of the charged residues (D, E, K, R, H, C, Y) and m*length of residue X > length of the pore in the first electrolytic cell. Let X be the negatively charged residue D. If X_1 is tethered to the membrane near the pore entrance in *cis*, then with V set to a large enough positive value (see Fig. 1(a)) the extended molecule $X_1 X_2 ... X_m R_1 R_2 ... R_n$ is drawn into the pore and remains fully stretched through the pore. Although the analyte portion $R_1 R_2 ... R_n$ is now fully exposed in *trans* a tRNA can attach itself only to R_n because this is where the carboxyl end is available for charging. In successive cycles R_n, R_{n-1}, ..., R_1 get attached to a corresponding tRNA, with each R being cleaved at the end of a cycle and the next R's carboxyl end getting exposed for charging. Thus tethering, with V set to a suitable value and polarity, ensures that the analyte remains close to the pore on the *trans* side so that charging and cleaving events occur in a known small neighborhood where they can be detected (optically) fairly easily.

2) Similarly RNA molecules in the electrolytic multicell can be tethered to the membrane in the respective *cis* chamber near the pore entrance on the *cis* side (see Fig. 1(b)). To ensure correct base pairing of the anticodon on a tRNA with the terminal codon of an RNA molecule in the multicell an RNA oligo should have k+3 bases where k * length of base ≈ length of pore such that only the terminal triplet is exposed in *trans*. Again with a suitable $V_i > 0$ the tethered RNA molecule, which carries a uniform negative charge, is drawn into the the i-th pore and remains fully stretched inside while only the terminal codon is exposed. However, the length of the pore places a rigid constraint on the value of k. One possible solution to this is to use a tethered DNA molecule with a dangling single strand 3-base extension consisting of only the triplet codon.

References

[1] J. M. Heather and B. Chain. "The sequence of sequencers: the history of sequencing DNA". *Genomics* **107**, 1–8, 2016.

[2] E. Pettersson, J. Lundeberg, and A. Ahmadian. "Generations of sequencing technologies". *Genomics* **93**, 105–11, 2009. doi:10.1016/j.ygeno.2008.10.003

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

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

[5] T. Rabilloud and C. Lelong. "Two-dimensional gel electrophoresis in proteomics: a tutorial". *J. Proteomics* **74**, 1829-1841, 2011.

[6] H. Bayley. "Nanopore sequencing: from imagination to reality". *Clin. Chem*. **61**, 25–31, 2015.

[7] D. Deamer, M. Akeson, and D. Branton. "Three decades of nanopore sequencing". *Nature Biotechnol*. **34**, 518–524, 2016.

[8] 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.

[9] E. Kennedy, Z. Dong, C. Tennant, and G. Timp. "Reading the primary structure of a protein with 0.07 nm^3 resolution using a subnanometre-diameter pore". *Nature Nanotechnol*. **11**, 968–976, 2016.

[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, .

[11] 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

[12] 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.

[13] J. Swaminathan, A. A. Boulgakov, and E. M. Marcotte. "A theoretical justification for single molecule peptide sequencing". *PLOS Comput. Biol*. **11**, e1004080, 2015.

[14] P. Boynton and M. Di Ventra. "Sequencing proteins with transverse ionic transport in nanochannels". *Sci. Rep*. **6**, 25232, 2016.

[15] J. Wilson, L. Sloman, Z. He, and A. Aksimentiev. "Graphene nanopores for protein sequencing". *Adv. Funct. Mater*. **26**, 4830–4838, 2016.

[16] 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.

[17] D. L. Nelson and M. M. Cox. *Lehninger's Principles of Biochemistry*, 4th edn., W H Freeman, 2005.

[18] B. Alberts, A. Johnson, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*, 6th edn., Garland Science, 2014.

[19] T. A. Steitz. "A structural understanding of the dynamic ribosome machine". *Nature Reviews* **9**, 242-253, 2008.

[20] A. J. Barrett, N. D. Rawlings, and J. F. Woessner. (eds.) *Handbook of Proteolytic Enzymes*, Academic Press, 1998.

[21] D. L. Crimmins, S. M. Mische, and N. D. Denslow. "Chemical cleavage of proteins on membranes". *Curr Protoc Protein Sci.*, 2001. doi: 10.1002/0471140864.ps1105 s19

[22] 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.

[23] J. Tullman, N. Callahan, B. Ellington, Z. Kelman, and J. P. Marino. "Engineering ClpS for selective and enhanced N-terminal amino acid binding". *Appl. Microbiol. and Biotechnol.* **103**, 2621–2633, 2019.

[24] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha. "Advances in single-molecule fluorescence methods for molecular biology". *Annu. Rev. Biochem.* **77**, 51–76, 2008.

[25] 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

[26] S. Carson and M. Wanunu. "Challenges in DNA motion control and sequence readout using nanopore devices". *Nanotech.* **26**, 074004, 2015.

[27] 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

[28] 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.

[29] J. Nivala. M. B. Marks, and M. Akeson. "Unfoldase-mediated protein translocation through an α-hemolysin nanopore". *Nature Biotechnol.*, **31**, 247–250, 2013.

[30] E. C. Yusko, B. R. Bruhn, O. Eggenberger, J. Houghtaling, R. C. Rollings, N. C. Walsh, S. Nandivada, M. Pindrus, A.R. Hall, D. Sept, J. Li, D. S. Kalonia, and M. Mayer. "Real-time shape approximation and fingerprinting of single proteins using a nanopore". *Nature Nanotech.*, 2016. doi: 10.1038/nnano.2016.267

[31] 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.* **7**, 10217, 2016. doi: 10.1038/ncomms10217

_____

Email: sampath_2068@yahoo.com