Self as the Core of Biological Feedback Systems: An Ideogrammatic Model of Memory and Recognition II (Proposed Experimental Designs for Validation)

Hyunho Shin (hyunho.shin@noctuabio.co.kr)

Abstract

This study proposes specific experimental approaches to validate the theory presented in "Self as the Core of Biological Feedback Systems: An Ideogrammatic Model of Memory and Recognition." First, it discusses the premise that information must possess at least a one-dimensional structure in physical form.

If a material responsible for memory exists within cells, it is highly likely to be linear proteins (e.g., talin, actin, microtubules) with one-dimensional structures that store information. To test this hypothesis, nanopore and Raman spectroscopy-based analytical techniques are designed. To apply nanopore technology derived from DNA sequencing to protein analysis, the study explores methods for either preserving the three-dimensional structure of proteins or converting them to their one-dimensional form during preprocessing, determining the physical characteristics of nanopores, and analyzing the resulting signals. Particularly, surface-enhanced Raman spectroscopy (SERS) is proposed as a method to analyze structural changes in proteins in real-time and evaluate their potential for information storage. This research presents a novel experimental approach to understand memory and information storage mechanisms at the protein level, providing concrete directions to validate the theoretical hypothesis.

Main

In "Self as the Core of Biological Feedback Systems: An Ideogrammatic Model of Memory and Recognition," I proposed the hypothesis that the memory systems of early lifeforms likely operated through a simple logographic mechanism—a direct connection between symbols and meanings—and that the distinction between "Self" and "Non-Self" serves as the starting point of biological activity. Furthermore, I suggested that consciousness begins with the recognition of "Self" and expanded this concept into a foundational Biological Feedback Operating System (FBOS) for living organisms.

However, my theory has focused primarily on philosophical and theoretical discussions rather than concrete biological mechanisms, resulting in a lack of experimental evidence and remaining largely at the level of hypothesis. Therefore, I have devised experimental methods to validate this theory and aim to discuss those methods in this paper.

Context and Structure of Information

Before discussing the experimental design for this theory, it is necessary to first examine how information, or biological information, is physically structured. It is well understood that DNA determines protein sequences and holds value as information. However, the individual bases of DNA (A, T, C, G) do not carry meaning as standalone elements. While each base can serve as a signaling molecule, the information each base carries can be considered zero-dimensional, akin to a point. When three bases come together, as in the case of "ATG" encoding methionine, the resulting sequence can be regarded as one-dimensional information formed by connecting those points. In other words, information must have at least a one-dimensional structure. High-level information can be represented in two or three dimensions, but it must minimally exist in a one-dimensional format.

Zero-dimensional information, such as individual numbers or bases like A, T, C, and G, constitutes information in itself but holds no meaning on its own. To function as meaningful information, these elements must exist within a specific context. For instance, one might argue that individual DNA bases contribute to cellular roles, such as ATP serving as the cell's primary energy source. While this is true, ATP as zero-dimensional information holds no inherent meaning. It only becomes meaningful within the context of being "a molecule that acts as the primary energy source for cells." Similarly, the number "1" has no intrinsic meaning but becomes significant when placed in a context like answering the question, "How many moons does Earth have?" In this way, all information must exist within at least a one-dimensional contextual framework to have meaning.

We know that DNA functions as encoded information to direct the synthesis of proteins, and these proteins' structures and expression timings play critical roles in numerous bodily functions. However, it is certain that DNA itself is not the medium for memory. While research methodologies for the brain have progressed from studying large-scale regions to finer scales, including the activation of neurons using tools like fMRI, the mechanism of how memory is stored remains unknown. Could this information reside within the cell itself? And if physical information must exist in at least a one-dimensional structure, it becomes plausible that linear proteins could serve this role. The question then becomes: Can we extract information from these linear proteins?

Technical Observation Limitations

Let us first examine the technical limitations of observation. To observe the vast expanse of the universe, we rely on instruments like the James Webb Space Telescope. On the other hand, to investigate subatomic structures and movements, we use particle accelerators. However, having technologies capable of observing the extremes of scale does not mean we can effectively observe everything in the intermediate range. For instance, cells and much smaller entities like viruses can be observed, but these observations are typically conducted on inactivated or fixed samples.

With advances in technology, it has only recently become possible to observe the movements of active cells. However, in the case of proteins, only purified and crystallized proteins can currently be visualized using methods such as X-ray crystallography or Cryo-EM. Observing the dynamic changes of active proteins in real time within cells remains a significant challenge. Thus, despite advancements, there is still a vast intermediate range of phenomena that elude direct observation.

We aim to investigate whether linear proteins contain any encoded information. The MeshCODE theory hypothesizes that talin stores binary information in a 13-bit format, while other theories suggest that actin, myosin, and microtubules may also store information. The commonality among these proteins is their linear structure. As previously discussed, information must possess at least a one-dimensional structure, which implies that physical information should inherently exist in a linear form. If proteins do encode information, linear proteins are the most likely candidates. Therefore, we must devise a method to extract information from these linear proteins.

On Sequencing Technology

Recent advancements in DNA sequencing technology have been remarkable, including next-generation sequencing (NGS), which uses short-read methods for high-speed, high-throughput sequencing, and Oxford Nanopore technology, capable of directly reading long DNA or RNA sequences from single molecules. I have speculated that nanopore technology could potentially be adapted to determine protein sequences, and I am aware that various research institutions are actively pursuing this line of investigation. Efforts to apply nanopore technology for protein sequencing are also underway. Before delving into this, however, it is worth discussing the challenges of protein sequence analysis.

Size:

Proteins are often considered larger macromolecules than DNA, and this is somewhat accurate when considering their three-dimensional structures. However, when comparing the primary structure of amino acid chains with that of nucleotide chains, DNA is generally larger. While nanopore technology has proven successful for DNA sequencing, amino acids are smaller molecules, which necessitates developing methods to detect such smaller structures.

Structure:

DNA maintains its stability through hydrogen bonds in its double-helix structure, which can be easily disrupted to access single-strand information. Proteins, however, are held together by various interactions, including hydrogen bonds, hydrophobic interactions, ionic bonds, van der Waals forces, and disulfide bonds. To access the amino acid information within peptide bonds, all these interactions must be disrupted, making the preprocessing of proteins significantly more complex.

Composition:

DNA is composed of only four bases (A, T, C, G), whereas proteins consist of approximately 20 different amino acids. This greatly increases the complexity of information when analyzing protein sequences. Current nanopore sequencing technology relies on detecting electrical signals as molecules pass through the pore, and distinguishing 20 types of amino acids presents a far more challenging problem than distinguishing just four DNA bases.

Next, let us consider the process of determining the sequence of DNA or proteins using nanopore technology, broken down into three main stages:

1. Preprocessing Stage

At this stage, the issues related to the size and structure of the molecules must be addressed. For DNA, the double helix must be unwound to produce single strands. For proteins, their three-dimensional structure must be unfolded into a linear, primary structure. In the case of proteins, this requires the removal of various bonds, including hydrogen bonds, disulfide bonds, hydrophobic interactions, and ionic bonds, making the preprocessing significantly more complex than for DNA.

2. Nanopore Translocation Stage

This stage involves determining how to pass the molecule through the nanopore and selecting an appropriate nanopore. For DNA, motor proteins are typically used to pull the strand through the pore. Proteins, however, present additional challenges due to their diverse sizes and chemical properties, necessitating different mechanisms for guiding and transporting the molecules through the nanopore.

3. Signal Detection and Interpretation Stage

As the molecule passes through the nanopore, signals generated by its movement must be interpreted to identify the individual bases (DNA) or amino acids (proteins). DNA can be distinguished by the four bases (A, T, C, G) through electrical signals. Proteins, on the other hand, have a much higher level of complexity due to the presence of approximately 20 different amino acids with varying sizes and chemical properties, requiring highly advanced signal interpretation technologies for accurate differentiation.

Using the Oxford Nanopore genetic sequencer as an example, in the preprocessing stage, motor proteins are attached to extracted DNA to convert it into a single strand and guide it through the nanopore. The prepared DNA passes through the nanopore, generating electrical signals that are analyzed to determine the nucleotide sequence. To apply this technology to protein sequencing, an efficient method for breaking down the three-dimensional structure of proteins into their primary structure during preprocessing is required. Additionally, electrical or optical signal interpretation technologies capable of accurately distinguishing signals from various amino acids as they pass through the nanopore are essential.

In a recent study by Keisuke Motone and colleagues, a method for decoding long protein sequences at the single-molecule level using nanopore-based technology was proposed. This research demonstrated the potential for high-resolution analysis of various proteoforms. The method employed the following features:

In the preprocessing stage, ClpX proteins were utilized. ClpX is an ATP-dependent protein unfoldase and translocase that converts the three-dimensional structure of proteins into their linear, primary structure. ClpX binds to the N-terminus of proteins and uses ATP to generate mechanical force to sequentially unfold the protein. Furthermore, as a translocase, ClpX

gradually moves the unfolded protein into the nanopore, approximately two amino acids at a time, to facilitate sequence reading. In the nanopore translocation stage, the CsgG nanopore was used. CsgG is a nanopore with dimensions and properties suitable for the stable translocation of proteins. In the signal detection stage, similar to the Oxford Nanopore approach, electrical signals generated as the protein passed through the nanopore were measured to decode the amino acid sequence.

Development of a Universal Protein Sequencer

However, this study faced challenges in distinguishing signals from certain amino acids, such as leucine and isoleucine. Additionally, the experiments primarily focused on synthetic protein sequences, leaving room for further investigation into naturally folded, complex proteins. Improvements in system processing speed and data collection efficiency are also necessary.
I aim to develop a more universally applicable protein sequencing method, building upon this research with some modifications.

1. Preprocessing Stage (Converting Tertiary to Primary Structure, Controlling Translocation Speed)

To convert proteins from their tertiary structure to a primary structure, a simpler approach commonly used in SDS-PAGE preprocessing was adopted. The buffer used in SDS-PAGE contains the following components, all of which facilitate the conversion of proteins from their tertiary to primary structure:

SDS (Sodium Dodecyl Sulfate): Disrupts hydrophobic interactions and imparts a negative charge to the protein.

DTT (Dithiothreitol) or β-Mercaptoethanol: Breaks disulfide bonds.

Heat treatment (95–100°C): Disrupts hydrogen bonds and weak non-covalent interactions.

Using this process, proteins can be converted into their primary structure. The preprocessed protein solution can then be loaded onto a long, thin acrylamide gel and subjected to electrophoresis, where the current can be adjusted to control the translocation speed. This enables proteins to be sequentially passed through the nanopore during the electrophoretic process, allowing for controlled analysis.

2. Selection of Nanopore Protein

When SDS-PAGE reagents are used for preprocessing, the SDS treatment denatures proteins, meaning the nanopore must remain stable in such buffer conditions. Alternatively, non-protein nanopores could be employed. However, precise nanopores capable of accommodating amino acids are not yet available. Thus, a protein nanopore that remains stable in the SDS buffer is required. Among potential candidates, aerolysin was selected due to its stability in high concentrations of SDS and its suitability in size for protein translocation.

3. Signal Analysis Method

When SDS-PAGE reagents are used for preprocessing, the SDS treatment causes proteins to carry a negative charge, making it impractical to distinguish 20 different amino acids purely through electrical signals. To address this, Raman spectroscopy was considered. Using Surface-Enhanced Raman Spectroscopy (SERS), aerolysin protein can be coated on a gold film to utilize the signal amplification effect of the gold surface. This approach allows for analyzing the protein sequence passing through the nanopore using SERS, thereby extracting amino acid information. However, the proposed nanopore-based protein sequencing system still requires

further validation. In particular, it must be confirmed whether SERS (Surface-Enhanced Raman Spectroscopy) has sufficient sensitivity to differentiate between amino acids in proteins.

The entire process is designed to be integrated into a single cell, enabling the sequential steps of converting proteins from tertiary to primary structure, passing them through the nanopore, and analyzing their Raman signals in a continuous manner. This system can be illustrated schematically as follows:

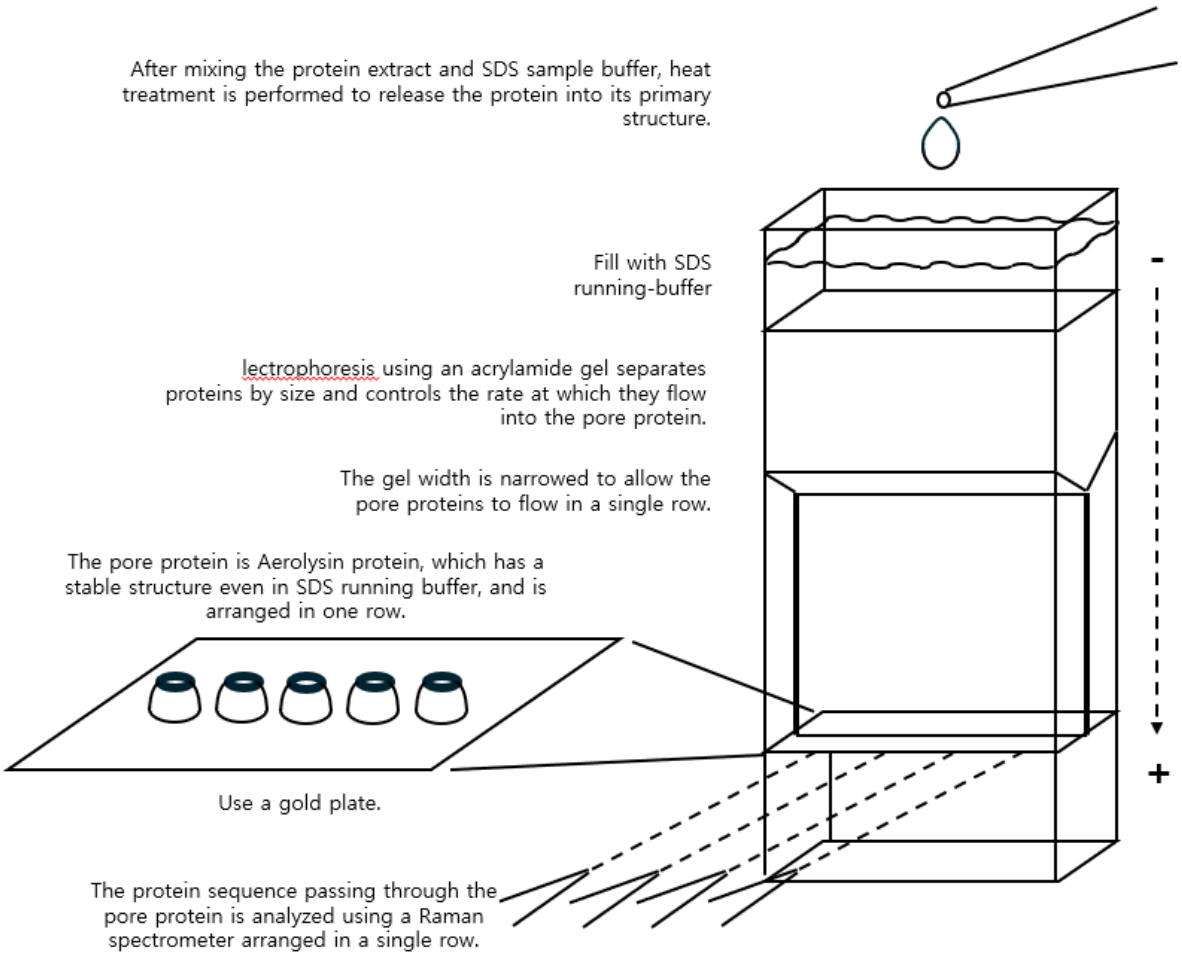

Figure 1. Development of a Protein Sequencer Using SDS-PAGE, Nanopore, and Raman Spectroscopy

Possibility of Information Storage in Linear Proteins and Proposed Validation Methods

So far, we have discussed the concept of a protein sequencer using nanopore technology. However, the ultimate goal is to determine whether memory can be stored as information within proteins. Specifically, we aim to extract information from linear proteins while maintaining their tertiary structure. For example, we need to investigate whether linear proteins such as talin, actin, myosin, and microtubules store binary information (1/0) or other types of data through their coiling and uncoiling. To achieve this, the proposed nanopore-based protein sequencer must be slightly modified.

1.Preprocessing Stage

To verify whether proteins in their tertiary structure can store linear information, the tertiary structure must remain intact during analysis, in contrast to converting proteins into their primary structure. After lysing the cells, the activity of proteases must be inhibited to preserve the structure of other proteins. Acrylamide gels, which were previously used to guide proteins through the nanopore, cannot be employed in this approach. An alternative method must be devised to guide proteins sequentially through the nanopore. Additionally, linear proteins must be separated from other types of proteins. For this purpose, microfluidics-based technology can be utilized, leveraging the unique alignment and movement patterns of long linear proteins in the fluid to achieve separation.

2.Selection of Nanopore

The nanopore must allow the passage of entire proteins, which requires a significantly larger pore size compared to those used for DNA or peptide sequencing. For instance, talin has a

molecular weight of approximately 400 kDa, and when fully extended in a linear configuration, it can reach a length of 70–100 nm, with a width of about 2–4 nm. The diameter of the nanopore must be at least 5–10 nm to accommodate the width of the protein without folding or bending. Ceramic, metallic, or silicon-based nanopores, which are typically fabricated with diameters ranging from 1 to 20 nm, are suitable for this purpose. Thus, there may be no immediate need to use protein-based nanopores.

3.Signal Analysis Method

Using Raman spectroscopy for signal detection is advantageous in this context. The focus is on detecting the degree of folding and unfolding of the protein to determine binary states (1/0) or other forms of information. This approach requires significantly less data than sequencing amino acids, making data collection and analysis more straightforward. Raman spectroscopy appears to be a practical and efficient solution for this purpose. This system can be illustrated schematically as follows:

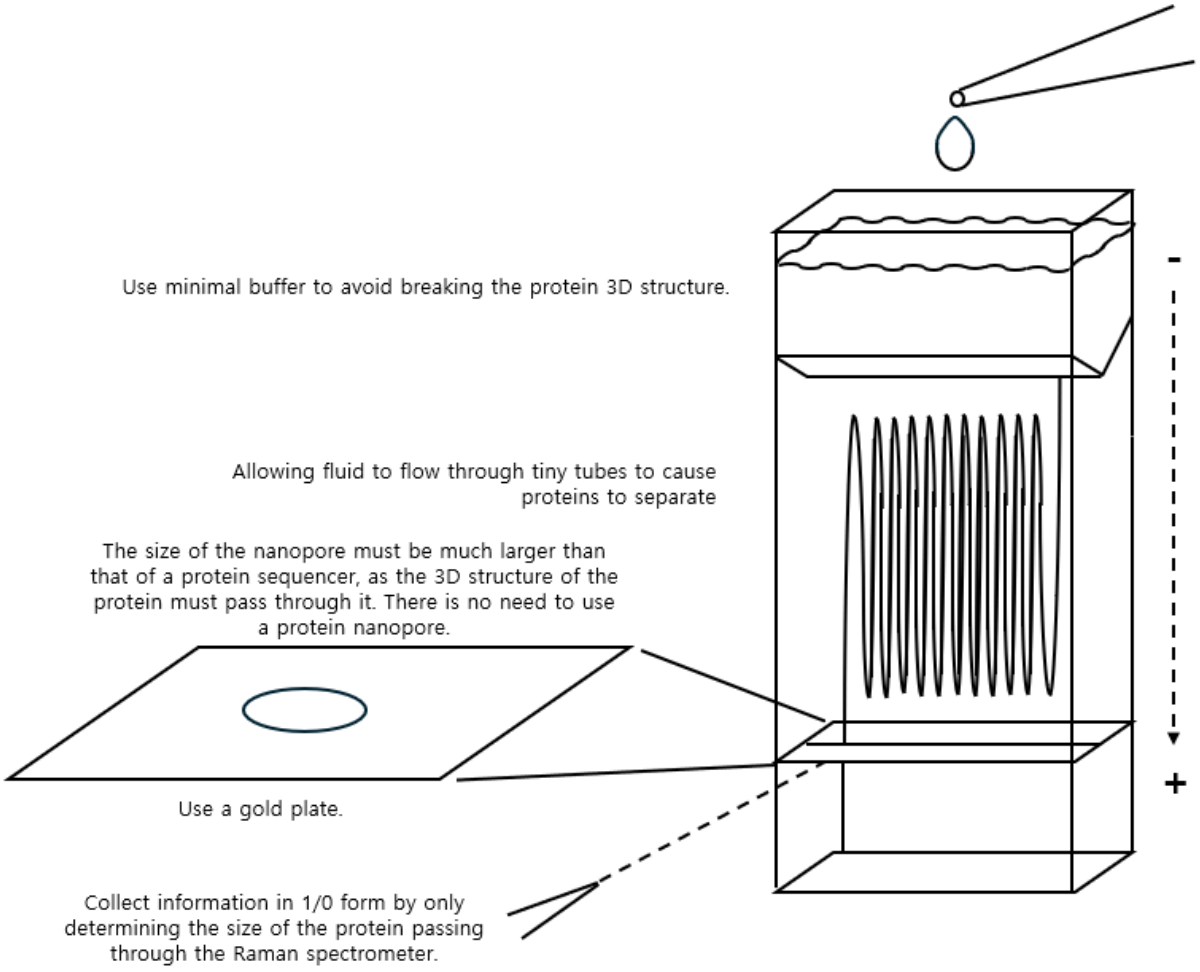

Figure 2. Modified Nanopore Protein Sequencer for Information Analysis in Linear Proteins

Mechanism Estimation Through Information in Talin or Microtubule Proteins

The proposed method can be applied to cultured neural cells or brain tissue to collect binary information encoded in linear proteins such as talin or microtubules using nanopore technology. The collected data can be analyzed to identify common information and overlapping patterns found across all proteins. Based on this extracted information, the binary data encoded in the proteins or the meaning of the obtained information can be inferred.

The analysis will focus on whether the information evolves hierarchically or follows specific patterns. Through this approach, we aim to uncover clues about memory and information storage mechanisms at the protein level and investigate whether the Fundamental Biological Operating System (FBOS), proposed in previous works as the framework underlying all memory, exists.

Reference

1. Self as the Core of Biological Feedback Systems: An Ideogrammatic Model of Memory and Recognition, *hyunho shin, viXra* (2024)

2. Talin in Mechanotransduction and Mechanomemory, *E. Klapholz et al.*

, Journal of Cell Science (2017)

3. The Mechanical Basis of Memory – the MeshCODE Theory, *B. T. Goult, Frontiers in Molecular Neuroscience* (2021)

4. Engineering Biological Nanopore Approaches towards Protein Sequencing, *Y. Wang et al.*

, Frontiers in Chemistry (2023)

5. Multi-pass, single-molecule nanopore reading of long protein strands, *Keisuke Motone et al.*

, Nature (2024)

6. Single-molecule protein sequencing with nanopores" J*ustas Ritmejeris et al.*

, Nature (2024)

7. Label-Free Optical Analysis of Biomolecules in Solid-State Nanopores, *M. P. Miguel et al.*

, ACS Photonics (2021)

8. SERS-Based Microfluidic Bioscreening Platform for Selective Detection of β-Amyloid Peptide

, S. Lee et al, Langmuir (2023)

9. Surface Enhanced Raman Spectroscopy for Single Molecule Protein Detection, *Lamyaa M. Almehmadi et al,Scientific Reports* (2019)

10. Protein Quantification and Imaging by Surface-Enhanced Raman Spectroscopy, *Hyunku Shin et al, Advanced Science* (2019)

11. Particle separation and sorting in microfluidic devices: A review, *P. Sajeesh et al, Microfluidics and Nanofluidics* (2014)