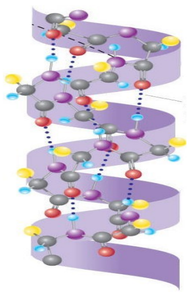

MACHINE LEARNING IN NEURODEGENERATIVE DISORDERS

Tribhuwan TIWARI , Roshan K BHATT, Ramya MANDLA and Prabir PATRA
Department of Biomedical Engineering, School of Engineering
University of Bridgeport, Bridgeport, CT

Abstract

Neurodegenerative disorders were responsible for 272,644 deaths in 2016 along in the US. The government spent $655 billion in 2020 for the direct and indirect medical costs from these diseases. Although experiments have been going on to find the cure for the neurodegenerative diseases, there has not been an efficient way till date to completely cure the diseases. Recently, there have been studies in understanding the cause of the diseases so as to plan for the early detection and treatment of diseases such Alzheimer's, Parkinson's disease and motor neuron diseases. Since proteins are the functional backbone of our body, understanding their formation and their relation to disease growth is a subject of interest. This is a study to link the application of a few methods developed in recent times to predict the presence of neurodegenerative diseases.

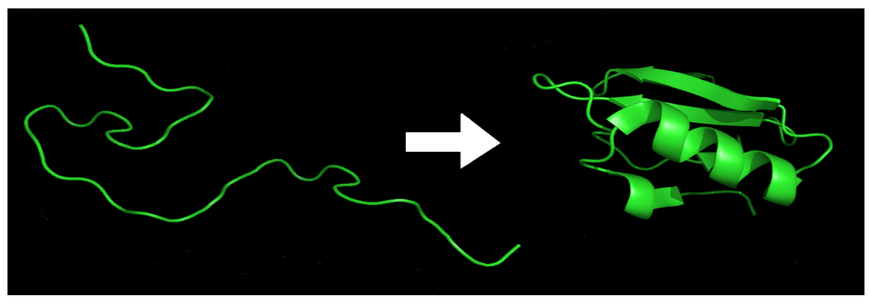

Fig 1.Protein as the linear chain and after folding.

Introduction

Proteins are the molecular machines that regulates and coordinates the vital cellular functions, which are composed of multiple number of amino acids. Protein is the biological polymer with conformational complexity. With an arsenal of 20 amino acids with very different chemical and structural properties, the unstructured heteropolymers synthesized by the ribosomes pack themselves into unique structures.

Proteins are the most versatile macromolecules, also known as building blocks of the body, which the mammalian cells typically compromise between 10,000 and 20,000 proteins. Protein synthesis, folding and degradation play pivotal role in the integrity and cellular health of the living organisms. The linear amino acids of synthesized polypeptide chain codes for the specific folded structure of protein. The biologically active conformation is often marginally stable under physiological conditions.

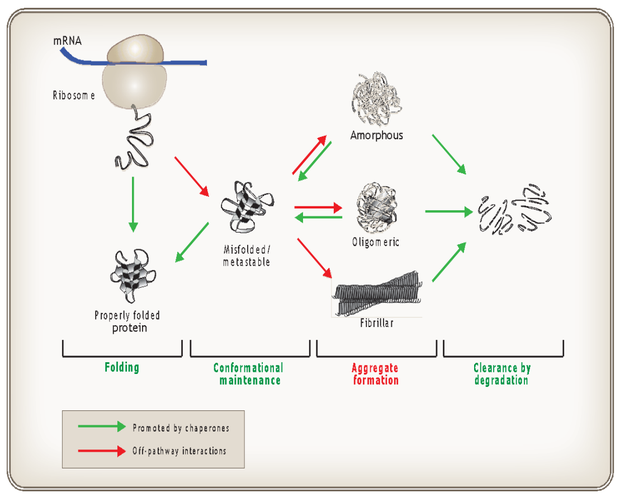

Fig 2. Protein Chain Folding and Mode of Action of Chaperones

Role of Chaperones in Protein Folding

Molecular chaperones are the proteinaceous structure that are highly regulated and coordinated which make the Protein Homeostasis Networks. The major roles by these molecules are the *foldase* activity leading to protein folding, *holdase* and *disaggregase* activity causing prevent or revert protein aggregation and take misfolded proteins for the degradation. Because of the upregulation in the stress conditions, this wide set of structurally diverse proteins is known as *heat-shock* or *stress* proteins (HSP).

Chaperones assist various proteins to fold efficiently and at biologically relevant rate. During translational process, chaperones proteins combine with the nascent polypeptide chain on the ribosome and cause folding to prevent (or reverse) misfolding and aggregation. Typically, these chaperones act by quickly shielding the hydrophobic amino acid residues that are exposed by proteins in their non-native conformations, but which are interred in the native state. They cooperate with machineries of protein degradation in a large protein homeostasis (or proteostasis) network.

Protein Misfolding and Diseases

Protein folding is an exquisitely regulated biological process through which a polypeptide is wrapped into a specific three-dimensional conformation that is critical for biological function. In the cell, this purely physics- and chemistry-driven process takes place in a highly complex biochemical environment whose interference is minimized by a tight proteostasis regulatory network that operates to maintain proteins in its adequate folding status.

Misfolding and aggregation of proteins are fundamental phenomena of protein biophysics leading to the development of several devastating neurodegenerative diseases, including Alzheimer's diseases (AD), Parkinson's diseases (PD), and Huntington's diseases (HD), amyotrophic lateral sclerosis, frontotemporal dementia, and the human prion diseases. The accumulation of abnormal protein aggregates detected as extracellular or intracellular deposits represents a common pathological signature for these diverse neurodegenerative disorders.

The first known protein-misfolding disease, indeed the first inherited human disease to have a known molecular mechanism, was **SICKLE CELL ANEMIA**. In this disorder, a single point mutation changes a glutamic acid in the β-globin chain of hemoglobin into a valine.

Fig 3. Aberrant Protein Folding Mechanism For Protein Folding Diseases

Protein Fold Prediction for Diseases Identification

On the basis of the evolutionary, physical and geometric constraints in protein structures, the AlphaFold Network greatly improves the accuracy of structure prediction by incorporating novel neural network architectures and training procedures. The AlphaFold network directly predicts the 3D coordinates of all heavy atoms for a given protein using the primary amino acid sequence and aligned sequences of homologues as inputs.

AlphaFold structures had a median backbone accuracy of 0.96 Å r.m.s.d.95 (Cα root-mean-square deviation at 95% residue coverage) (95% confidence interval = 0.85–1.16 Å). In addition to very accurate domain structures, AlphaFold can produce highly accurate side chains when the backbone is highly accurate and considerably improves over template-based methods even when strong templates are available.

Evoformer blocks are new mechanisms to exchange information within the MSA and pair representations that enable direct reasoning about the spatial and evolutionary relationships.The main principle of Evoformer is to view the prediction of protein structures as a graph inference problem in 3D space in which the edges of the graph are defined by residues in proximity. The structure module is the next structure after Evoformer which takes the representation and builds three-dimensional structure for the protein, including the position of the side chains. Together with a plethora of deep learning tricks, the model produces breathtakingly accurate predictions.

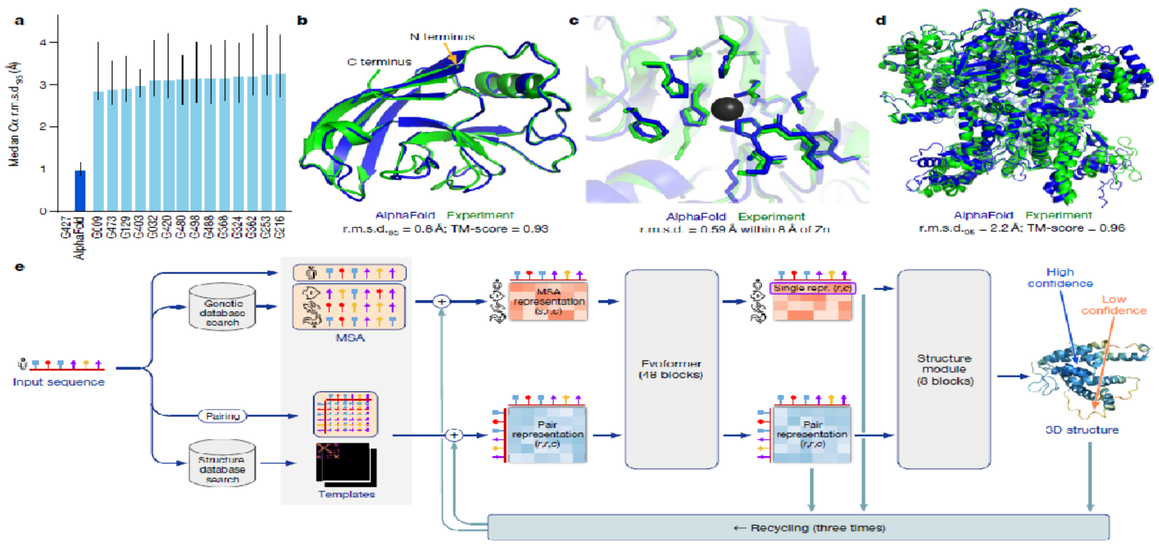

Fig 4. Prediction of Protein Folding and its Structure by AlphaFold 2

Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases

Disease	Genetic causes	Function
Alzheimer's disease	APP	Gives rise to Aβ, the primary component of senile plaques
Parkinson's disease	PS1 and PS2	A component of γ-secretase, which cleaves APP to yield Aβ
Parkinson's disease	α-Synuclein	The primary component of Lewy bodies
Parkinson's disease	Parkin	A ubiquitin E3 ligase
Parkinson's disease	DJ-1	Protects the cell against oxidant-induced cell death
Parkinson's disease	PINK1	A kinase localized to mitochondria. Function unknown. Seems to protect against cell death
Parkinson's disease	LRRK2	A kinase. Function unknown
Parkinson's disease	HTRA2	A serine protease in the mitochondrial intermembrane space. Degrades denatured proteins within mitochondria. Degrades inhibitor of apoptosis proteins and promotes apoptosis if released into the cytosol
Amyotrophic lateral sclerosis	SOD1	Converts superoxide to hydrogen peroxide. Disease-causing mutations seem to confer a toxic gain of function
Huntington's disease	Huntingtin	Function unknown. Disease-associated mutations produce expanded polyglutamine repeats

Lin, M. T. & Beal, M. F. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. Nature 443, 787–795 (2006) doi:10.1038/nature05292.

Conclusion

Protein folding and misfolding characterizes the basic role in the development of healthy cells and individuals. Neurodegenerative disorders pathologically express the accumulation of aggregation of abnormal protein in the form of extracellular or intracellular deposits. To combat the complication of neurodegenerative disorders, the study of chaperones alongside the associated protein is of prime importance. With the advent of the AlphaFold, the prediction accuracy has been higher compared to all other approaches. AlphaFold has developed the innovative machine learning technique which includes the physical and biological characteristics of protein like structure, multisequence alignments with the design of deep learning problem. Thus, Alphafold has indicated that protein folding can be unveiled to demonstrate the realms of different neurodegenerative disorders.

References

1.Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

2.Lyubchenko, Y. L., Kim, B. H., Krasnoslobodtsev, A. V., & Yu, J. (2010). Nanoimaging for protein misfolding diseases. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, *2*(5), 526–543. https://doi.org/10.1002/wnan.102

3.Naganathan, A. N. (2020). Protein folding: how, why, and beyond. In *Protein Homeostasis Diseases*. Elsevier Inc. https://doi.org/10.1016/b978-0-12-819132-3.00001-4

4.Sun, P. D., Foster, C. E., & Boyington, J. C. (2004). Overview of protein structural and functional folds. *Current Protocols in Protein Science / Editorial Board, John E. Coligan ... [et Al.]*, *Chapter 17*, 1–189. https://doi.org/10.1002/0471140864.ps1701s35

5.Thorpe, K. E., & Levey, A. I. (2021). *OF NEURODEGENERATIVE DISEASE Literature Review Summary May 2021 Alzheimer's disease and related dementias*. *May*

6.Valastyan, J. S., & Lindquist, S. (2014). Mechanisms of protein-folding diseases at a glance. *DMM Disease Models and Mechanisms*, *7*(1), 9–14. https://doi.org/10.1242/dmm.013474