

*Ministry of External Affairs  
NEST Division*

*Artificial Intelligence for expediting discovery biology &  
developing new strategic therapies.*

***Introduction***

Therapeutic proteins are either natural or engineered protein molecules developed for medicinal use. Therapeutic proteins are used for the treatment of various diseases like Cancer, Autoimmune disorders, Aging associated disorders like Macular degeneration and viral infections. Proteins are important macromolecules made of long chains of 20 natural aminoacids. Based on the sequence and length of the chains of the aminoacids, the proteins organize themselves into 3-dimensional structures which in turn determines the function. Understanding the structure of the proteins is a key bottleneck in development of effective strategies to combat human diseases.

2. Typically solving the protein structure involves experimental biophysical techniques like nuclear magnetic resonance (NMR), X-ray crystallography and Cryo-Electron microscopy. Techniques mentioned are not only laborious but also time consuming.

3. Timing is crucial especially during pandemics and effective therapies are required in a short span of time. Ongoing COVID-19 pandemic brings us to a self-realization that we be adept in cutting edge scientific technologies for pacing out new discoveries to combat diseases at a nascent stage.

4. With advent of machine learning and Artificial Intelligence (AI), new algorithms like Alphafold and Rosetta commons are being developed that expedite the solving of protein structure with very high accuracy. The new AI algorithms would help in expanding our knowledge of basic biology and also create new avenues for targeted and effective therapies.

## ***Alphafold***

6. Traditional *In silico* structure prediction of novel proteins depends on the homology of the amino acid sequence to the previously experimentally solved proteins.

7. Free modelling/ Ab initio modelling (solving the structure from the basic amino acid sequence) has been a challenge and relies on modelling parts of the larger protein and stitching them to determine the consensus structure. In Free modelling approach the structure hypothesis is regularly modified and the structures predicted are low potential structures.

8. [Alphafold](#) algorithm utilizes deep learning approach, where it takes into consideration the twisting of the backbone and distance between the side chains residues of the aminoacids to predict the structure of the protein. The neural network of Alphafold is trained to identify the patterns in structures (about 1,50,000) deposited in Protein Data Bank.

9. CASP (Critical Assessment of protein Structure Prediction) is a bi annual competition to predict the structure of proteins computationally and compare head to head to experimentally solved structures. Alphafold in the latest CASP competition not only outperformed the other competitors in predicting the protein structure but attained near accuracy comparable to experimental results.

10. Scientific Journal, Nature has called the study as “gargantuan leap in solving biology’s grandest challenge – determining protein structure from its amino-acid sequence”. Scientists also herald Alphafold as a ‘gamechanger and boon to life sciences’, since it would accelerate our understanding of building blocks of cell and thus aid advanced drug discoveries.

11. Alphafold algorithm is developed by DeepMind (subsidiary of Google) and is not open source but have plans to make it available.

## ***Rosetta commons***

12. Rosetta commons is a hub of protein modelling software that has algorithms to predict structure. The other goals of the software are i) Understanding macromolecular interactions ii) designing custom molecules iii) developing efficient ways to search confirmation for docking studies.

13. One of the recent applications of Rosetta software was to design novel '[mini-proteins](#)' (not occurring in nature) against SARS-CoV-2 viral Spike protein. Spike protein is the main antigenic site on SARS-CoV-2 virus that binds to human Angiotensin-converting enzyme 2 (ACE2) on cell surface and gains entry into our systems. The designed mini-proteins bind to the viral Spike protein and antigenic site becomes unavailable for binding to human ACE2, thus preventing infections. The advantage of mini-proteins is they are small, stable and can be delivered via the nasal spray. Two of the best mini-proteins have been found to be effective in picomolar concentrations and are currently in clinical trials.

14. Rosetta is an opensource platform for educational research purposes.

## ***Suggestions***

15. It is apparent that a multi-layered approach is necessary for fighting pandemics. We require development of vaccines, small molecule drugs, antibiotics and biologics to treat new and emerging diseases.

16. Obtaining the genomic sequence of the circulating viruses or other pathogens is fairly easy with currently available DNA sequencing technologies. We can identify the Open reading frames (ORF's)/Protein coding regions of the genomic sequence, which can further be translated to the amino-acid sequence of the proteins. AI technologies mentioned above can be used to determine the structure of the proteins and thus help us understand the antigenic sites and drug targetable regions in the viruses.

17. Alphafold and Rosetta programs can be used to develop targeted strategies for diseases like Dengue fever, Malaria, Chikungunya, Tuberculosis that are of high

burden in India. These new AI technologies can be applied for research into emerging viral infections like Ebola, Nipah, Lassa which have very high mortality rates in comparison to recent Covid-19 pandemic.

18. Alphafold and Rosetta prove the power of inter-disciplinary approaches leading to new Innovations. Collaborations with like minds would help in solving complex questions and addressing the medical challenges of the 21<sup>st</sup> century.

19. Biomedical granting agencies can promote AI based research by inviting special calls for granting. Ind-CEPI's (Coalition for Epidemic Preparedness Innovations) mission is to develop vaccines to current and emerging virus infections in India.