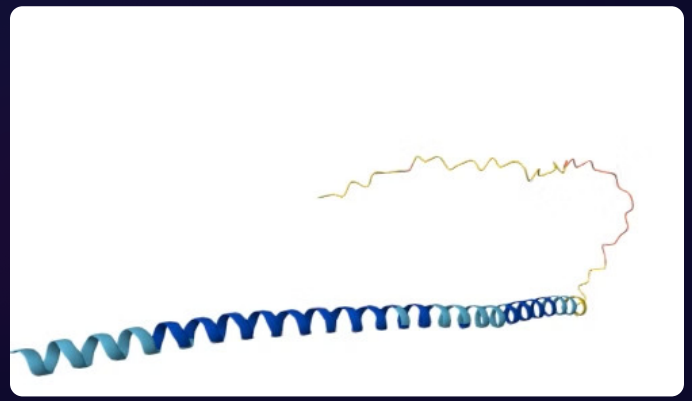
Hybrid Quantum-AI Approach for Protein Folding Problem and Neurodegenerative Disease Prediction

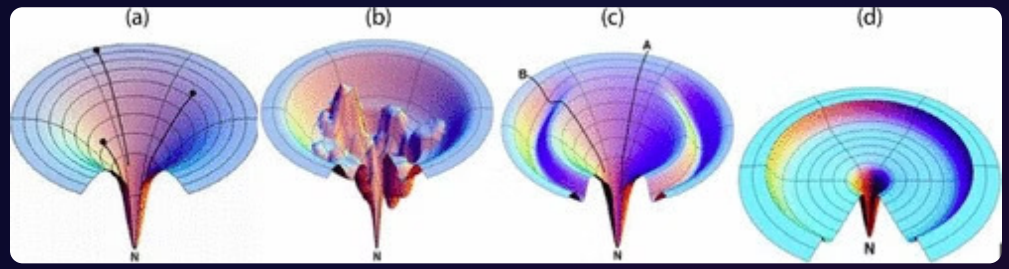
Hackathon Project Presentation



Our team is excited to present a groundbreaking approach that combines the power of quantum computing and artificial intelligence to tackle the challenges of protein folding problem and neurodegenerative disease prediction. Through innovative hybrid modeling, we aim to push the boundaries of what's possible in these critical fields of research.

Team Members **:** [Bhavik, Amy, Vishal]

The Challenge: Protein Folding and Neurodegenerative Diseases



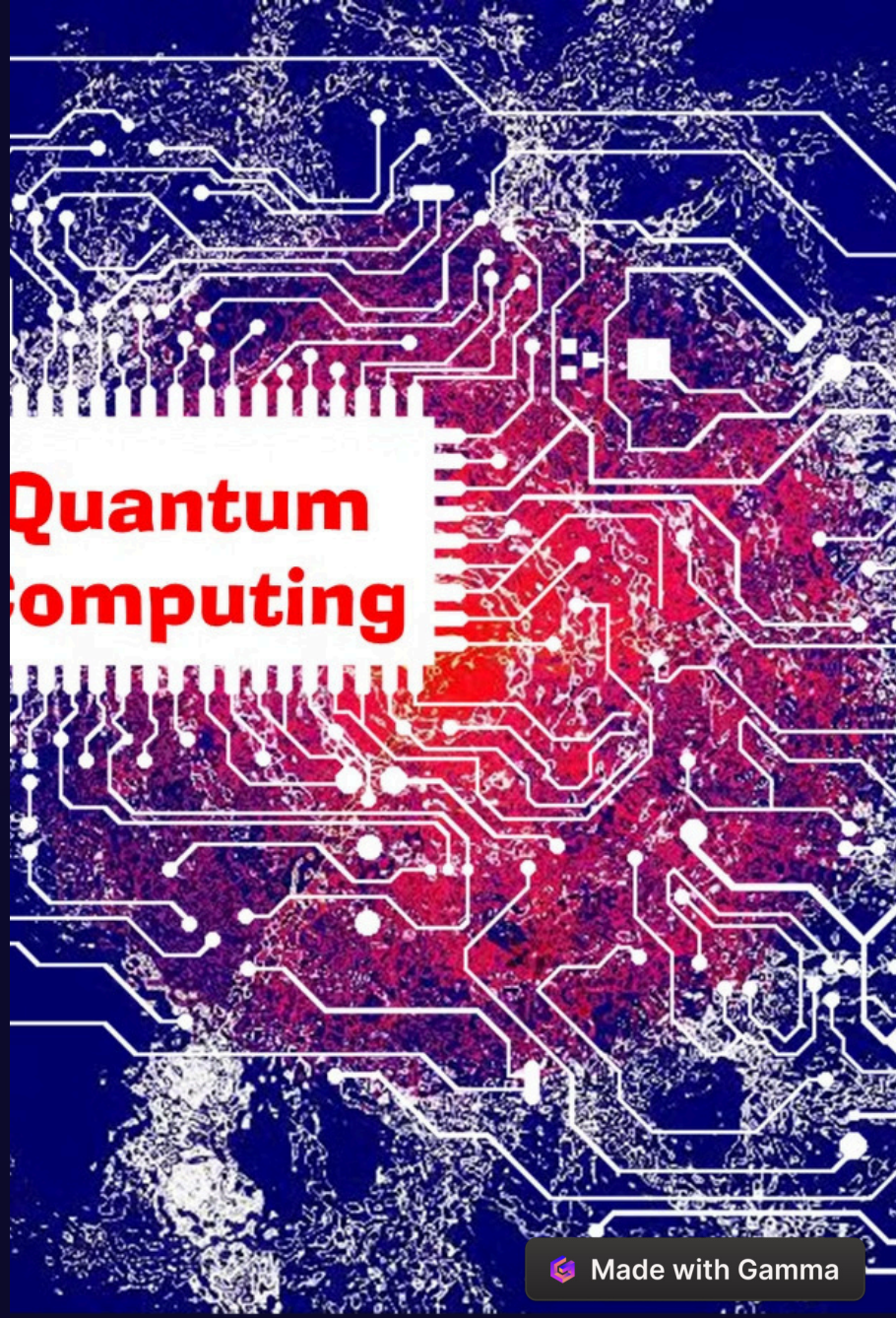
- **Rugged Energy Landscape**: Proteins must navigate many local folding states to reach the correct structure, which is computationally challenging.

The Complexity of Protein Folding

- **Increasing Sequence Length**: creates exponentially more possible structures.
- **Mutation**: can significantly affect its folding and structure.
- **Intrinsic Disorder (IDRs)**: Disordered regions lack stable structures, complicating prediction.
- **MSA (Multiple Sequence Alignment)**: Generating accurate alignments becomes harder with sequence diversity.

Quantum Computing Overview

Protein folding is a complex biological process where a protein's amino acid sequence determines its three-dimensional structure. Solving the protein folding problem (PSP) has been a longstanding challenge due to the vast conformational space proteins can adopt. Classical algorithms often struggle with the computational complexity of simulating these interactions. Quantum computing offers a new approach, leveraging quantum mechanics to simulate molecular interactions more efficiently. Quantum algorithms could potentially navigate protein folding's complex conformational space, improving prediction accuracy and speeding up calculations. While quantum hardware and algorithms are advancing, determining the types of PSP problems that can benefit from quantum advantage remains an open question. This research explores how quantum computing might enhance protein folding solutions.



Project Goals

1 Protein folding with Variational Quantum Eigenolver (VQE)

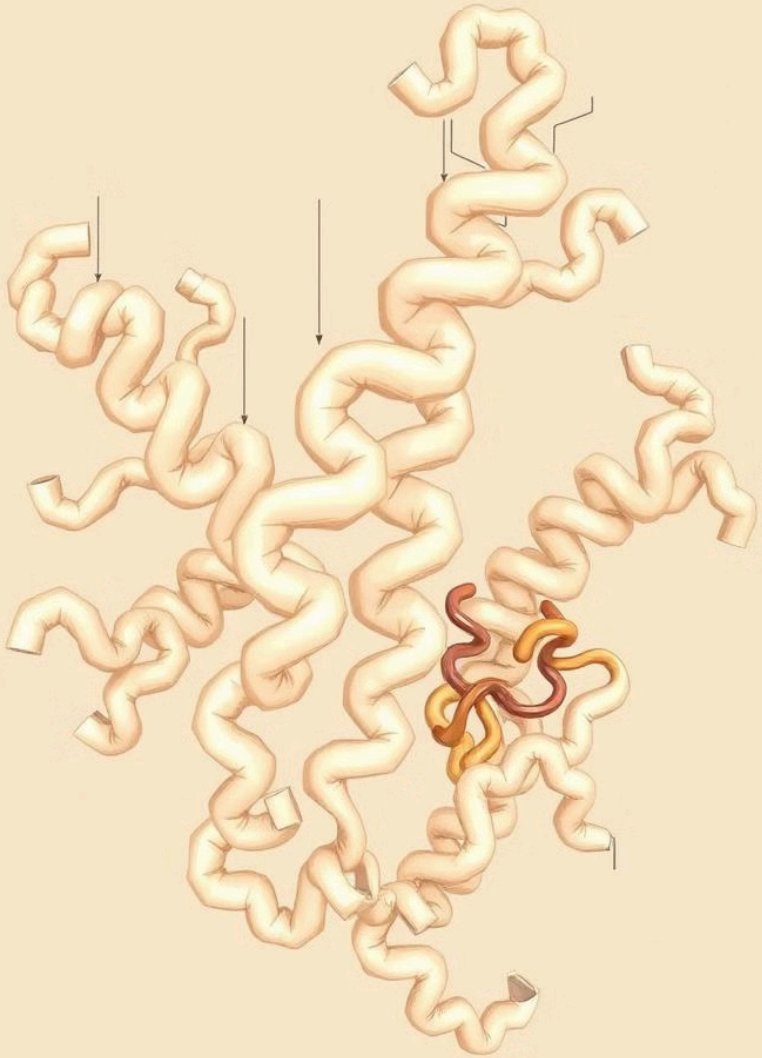
calculating the Hamiltonian for a protein sequence, incorporating various factors such as sequence length, mutation penalty, intrinsic disorder region (IDR) penalty, and conservation score. Then using Gradient Descent optimized VQE to find optimized parameters corresponding to ground energy state of that hamiltonian.

2 Hybrid Quantum AI model

A hybrid Quantum-AI model combines a quantum circuit for feature embedding with a classical neural network. The quantum layer uses these embedding and an ansatz to generate expectation values, which are fed into a classical neural network for binary classification.

3 Enhanced Drug Discovery

Insights from quantum parameters and structural features that could represent misfolding and ligand-protein interaction, which is essential for understanding protein aggregation and toxicity that leads to diseases. This can enable us to find drugs that can inhibit these aggregations.



Our Hybrid Model

Combining AI and Quantum Strengths

Calculating the **Hamiltonian** of a protein sequence, incorporating factors such as sequence length, mutation penalties, IDR penalties, and conservation scores. The Hamiltonian serves as a model for the protein's energy landscape. Using **Gradient Descent-based VQE**, the model optimizes the Hamiltonian to find the ground state energy, which represents the protein's most stable conformation.

The quantum layer then applies **Embedding** and an ansatz to embed these features into quantum states, generating expectation values. These quantum outputs are passed into a **classical neural network**, which processes them for binary classification, potentially identifying behaviors like misfolding or protein-ligand interactions. This hybrid approach effectively combines quantum computing for optimization with classical AI for classification, offering a powerful tool for understanding protein dynamics and improving drug discovery efforts.

Data Preparation

Protein Sequence Retrieval

Protein sequences were fetched using the UniProt API. Target proteins included alpha-synuclein, tau, and amyloid beta. These are primarily linked to neurodegenerative diseases.

IDRs

using IUPred2A tool for intrinsically disordered proteins (IDRs) and the disorder scores are calculated.

Multiple Sequence Alignment (MSA)

using MSA to capture evolutionary relationships and conserved structural motifs and domains in protein for predicting protein structures or functions based on homologous sequences.

Identifying Disordered Regions

Intrinsic Disorder Regions (IDRs) are segments of proteins that lack a stable 3D structure and can adopt multiple conformations based on their environment. These regions play crucial roles in functions like protein-protein interactions, signaling, and cellular regulation. Proteins with IDRs, known as **Intrinsically Disordered Proteins (IDPs)**, are highly flexible and involved in critical processes such as molecular recognition.

Tools like **IUPred2A** predict IDRs based on amino acid sequences by assigning **disorder scores**, with higher scores indicating a greater likelihood of disorder. Understanding IDRs is vital for studying disease mechanisms, especially in neurodegenerative diseases where misfolding and aggregation of IDPs, like **Alpha-synuclein, Tau,** and **Amyloid-beta**, contribute to pathology.

Conserving Structural Insights

Alignment of sequences for evolutionary insights

Multiple Sequence Alignment (MSA) employed to align protein sequences. This process reveals evolutionary relationships and conserved regions within the protein family, providing crucial insights into structure and function.

Calculation of conservation scores

Conservation scores calculated for each amino acid position in the alignment, quantifying the degree of sequence similarity across different species. High conservation scores suggest functionally important residues crucial for maintaining protein structure and stability.

Visual representation: Heatmap of MSA

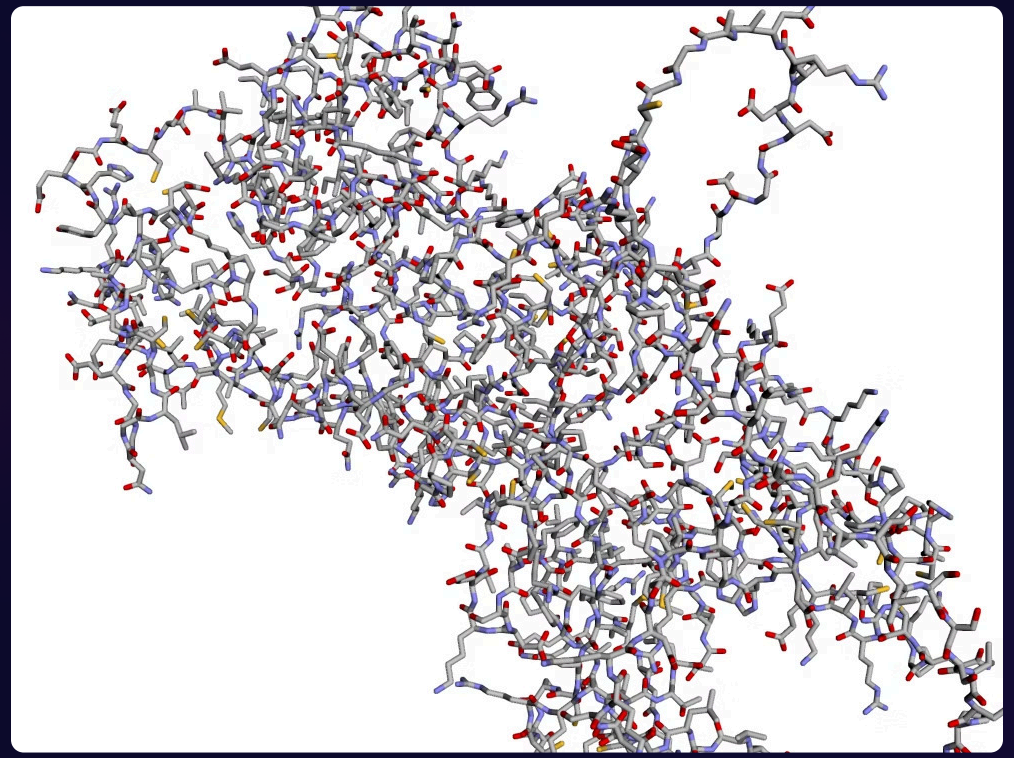
A heatmap visualization generated to represent the MSA and conservation scores. This provides a clear visual representation of conserved regions, highlighting areas of structural importance across the aligned protein sequences.

Quantum Hamiltonian for Stability

The stability of a protein conformation is quantified using a quantum Hamiltonian (H). A lower Hamiltonian value indicates a more stable conformation. This stable energy state is found using VQE algorithm.

This Hamiltonian incorporates several key factors:

- **Sequence Length:** Longer sequences generally have higher Hamiltonian values due to increased interaction complexity.
- **Mutation Penalty:** Mutations that destabilize the protein structure increase the Hamiltonian value.
- **IDR Penalty:** The presence of intrinsically disordered regions (IDRs) contributes to the Hamiltonian, reflecting their potential for instability.
- **MSA and Conservation scores:** average conservation score from the provided list of conservation scores for each position in the MSA. Higher conservation scores indicate higher conservation at those positions across the aligned sequences.



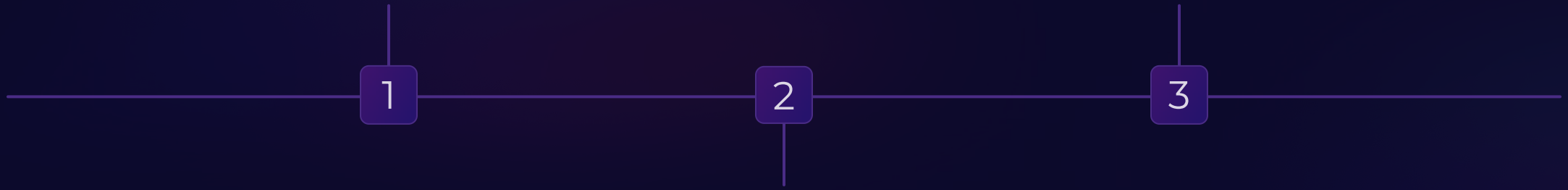
Integrated Prediction Workflow

Quantum Layer

The quantum layer of the hybrid model handles the complex, many-body interactions inherent in protein systems, leveraging the power of quantum computing.

Hybrid Prediction

The integrated workflow seamlessly integrates quantum and AI approaches, harnessing the strengths of both to deliver accurate and efficient protein folding predictions. Simulating misfoldings and ligand interactions.



Neural Network Integration

A neural network combines the outputs from the quantum layer with classical machine learning features to provide a comprehensive prediction of protein stability and structure.

Results :

Implementation of VQE with PennyLane

We implemented the Variational Quantum Eigensolver (VQE) algorithm using the PennyLane quantum machine learning library. PennyLane's user-friendly interface allowed for efficient development and testing of the VQE algorithm in our protein stability analysis. The VQE algorithm is used to optimize the parameters of a quantum circuit for the Hamiltonians of proteins (SNCA, Tau, and Abeta). H_snca, H_tau, and H_abeta are Hamiltonians that describe the energy functions for the respective proteins using Pauli operators (PauliZ). The ansatz function creates a variational quantum circuit where each qubit is rotated by RY and RZ gates, followed by CNOT gates between neighboring qubits. This circuit is parameterized by params. The VQE optimization loop iterates to minimize the energy by adjusting the parameters using gradient descent.

Integration with neural network for prediction

The classical layer is a simple feed-forward neural network (nn.Linear) that takes the quantum outputs as input and predicts a binary classification output using a sigmoid function.

Conclusion

This presentation detailed a novel hybrid quantum-AI model for simulating protein folding and predicting neurodegenerative diseases. Our projects builds on to the existing research and extending this to open new avenues for drug discovery, personalized medicine, and a deeper understanding of complex biological processes related to protein misfolding. The hybrid model of integration of AI and quantum computing enables us to leverage the power of quantum computing while addressing challenges of quantum computation hardware resources and challenges. This also addresses the scalability challenges.

Future Scope

Integrating experimental data from techniques like X-ray crystallography or NMR spectroscopy to improve the accuracy of the Hamiltonian models and better predict protein stability and misfolding.

The VQE optimization could be extended to consider multiple objectives, such as stability, binding affinity, and toxicity, simultaneously. This would provide a more holistic view of protein behavior and its interaction with potential ligands, making the model more relevant for drug discovery.

Predict how proteins interact with potential drugs. By optimizing drug candidates at the quantum level, these hybrid models can speed up drug screening, enhance binding affinity predictions, and improve personalized medicine approaches. This approach can potentially lead to faster, more effective treatments for complex diseases by targeting the underlying protein dysfunctions.

Thank You