Multi-Modality Approach for Predicting the Impact of Protein Mutations

Reza Marzban [1] **Alireza Kheirandish** [1] **Amirali Aghazadeh** [1]

Abstract

Predicting the impact of protein mutations on fitness and stability is a critical challenge in protein engineering and biological research. In this study, we propose a multi-modality framework that integrates sequence-based embeddings from ESM2 and structure-based embeddings from AlphaFold 3. To evaluate the utility of these combined embeddings, we implement two complementary model architectures: Cross-Attention and Graph Attention Networks (GAT). Both approaches demonstrate the capability to integrate sequence and structural data effectively, achieving validation accuracies of **83.5%** and **84%**, respectively. These results underscore the versatility of our framework in modeling protein mutation impacts, offering robust tools for future applications in protein engineering.

1. Keywords

Multi-Modal, Protein Mutations fitness predictor, Graph Attention Network, Cross Attention model, Deep Mutational Scanning

2. Introduction

Predicting the effects of mutations on protein fitness is a complex and pivotal challenge with far-reaching implications in drug design, synthetic biology, and understanding disease mechanisms (Cheng et al., 2024). Mutations in protein sequences can alter folding, stability, and function, potentially leading to enzyme deficiencies, human diseases, or viral escape (Anstrom et al., 2005; Cocco et al., 2024). Consequently, accurately modeling mutation effects is crucial for designing protein variants with enhanced or novel functionalities, offering transformative possibilities in biotechnology and biomedicine.

Traditional approaches to mutation effect prediction have primarily focused on leveraging either sequence-based or structure-based data. Sequence-based models capture evolutionary relationships, such as those encoded by ESM2 embeddings (Zeng et al., 2024), while structure-based approaches like AlphaFold embeddings provide insights into protein folding and structural conformations (Akdel et al., 2022). However, relying solely on one modality can be limiting. For instance, sequence-only methods may fail to capture certain long-range structural interactions, while structure-only methods may miss evolutionary signals embedded in the sequence data. These long-range structural interactions are often essential for understanding protein function and cannot be fully inferred from sequences alone (Tang & Kaneko, 2020).

Efforts to integrate both sequence and structure information are emerging. For example, multimodal models, such as the Protein Mutational Effect Predictor (ProMEP), have demonstrated state-of-the-art performance by combining deep representation learning with sequence and structure contexts. ProMEP's success in capturing intricate mutational impacts underscores the potential of multimodal frameworks to enhance both predictive accuracy and computational efficiency (Cheng et al., 2024). Yet, the accurate prediction of mutation effects, particularly for less-studied or de novo-designed proteins, remains a fundamental challenge due to the complexity of residue interactions and mutational epistasis (Anstrom et al., 2005).

In this work, we address these limitations by proposing a novel multi-modality framework that integrates embeddings from ESM2 and AlphaFold 3. Our approach evaluates two complementary architectures—Cross-Attention and Graph Attention Networks (GAT)—to combine sequence and structural information effectively. Our key contributions are outlined below.

- Development of a multi-modality framework that integrates sequence and structural embeddings for mutation impact prediction.
- Implementation of two model architectures—Cross-Attention and GAT—each tailored to leverage the unique strengths of multimodal embeddings.
- Empirical validation of both models on diverse protein

[1] School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, USA. Correspondence to: Reza Marzban <mmarzban3@gatech.edu>, Alireza Kheirandish <akheirandish3@gatech.edu>.

datasets, showcasing their utility for predicting protein mutation impacts.

By incorporating insights from state-of-the-art methods like ProMEP and leveraging advanced attention-based architectures, our study aims to push the boundaries of mutation effect prediction and contribute robust tools for protein engineering and biological research.

3. Methodology

3.1. Dataset

We selected three datasets from ProteinGym (Notin et al., 2024) to represent diverse protein types and functions:

- **LISMN**: Human mitochondrial proteins involved in energy production (Seuma et al., 2022).
- **ARGR_ECOLI**: *E. coli* arginine biosynthesis proteins crucial for amino acid synthesis (Tsuboyama et al., 2023).
- **BBC1_YEAST**: Yeast actin-binding proteins important for cytoskeletal structure (Tsuboyama et al., 2023).

We chose 301 samples from each dataset annotated with Deep Mutational Scanning (DMS) scores, reflecting the fitness impact of mutations (single or multiple amino acids). The binary DMS scores already exist in the dataset and were created using a threshold specific to each wild-type protein. If the mutation fitness is higher than the threshold, the mutation is considered effective or positive; otherwise, it is deemed deleterious. This data facilitates a classification task to predict the functional impact of mutations.

3.2. Embedding Extraction

To capture both structural and sequence information, we extracted embeddings from two state-of-the-art models:

AlphaFold 3 Structural Embeddings: Using AlphaFold 3 (Abramson et al., 2024), we generated 64-dimensional embeddings for each protein, encapsulating spatial and folding information.

ESM2 Sequence Embeddings: Leveraging the ESM2 model (Rao et al., 2020), we obtained 1280-dimensional embeddings that reflect the protein's evolutionary history and sequence relationships.

To address the dimensionality difference, we applied z-score normalization to both embeddings to ensure compatibility for integration into our models. No dimensionality reduction was applied, preserving the richness of the original embeddings.

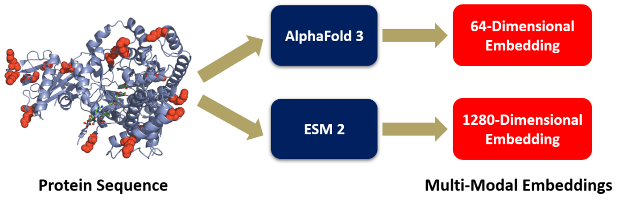


Figure 1. Ribbon diagram of *E. coli*(Anstrom et al., 2005) malate synthase G, highlighting mutated lysine residues (in red) near the active site with Coenzyme A and pyruvate. This diagram illustrates the regions where mutations occur, providing context for the embedding extraction process. By processing the protein sequences through AlphaFold 3 and ESM2, we obtain structural and sequence embeddings, respectively, which are then combined for DMS score prediction.

As shown in Figure 1, mutations often occur near active sites, affecting protein function. Capturing both sequence and structural information is crucial for accurate predictions.

3.3. Model Architectures

3.3.1. Cross-Attention Model

Our Cross-Attention Model consists of 2 layers with 4 attention heads (Hou et al., 2019). It projects the ESM2 and AlphaFold embeddings into a shared latent space, generating queries, keys, and values for the attention mechanism. The cross-attention layers allow the model to learn relationships between sequence and structural features by attending to one modality while considering information from the other. A sequence length of 8 was used to leverage cross-attention by creating patches to capture relationships between two embeddings. The concatenated attention outputs are then passed through a fully connected layer and a softmax classifier for prediction. The image in Figure 2 illustrates how the model integrates the two embeddings

3.3.2. Graph Attention Network (GAT)

The GAT model combines the embeddings by concatenation, forming a feature-rich input for graph processing. It consists of 4 GAT layers, each employing multi-head attention mechanisms to capture complex relationships between nodes (representing features from both embeddings). The GAT effectively models non-local interactions and leverages the graph structure inherent in protein data. Pooling mechanisms—including global attention, max, and mean pooling—aggregate the learned features, which are then fed into a classifier.

Figure 3 illustrates the GAT architecture, showcasing how it processes the combined embeddings.

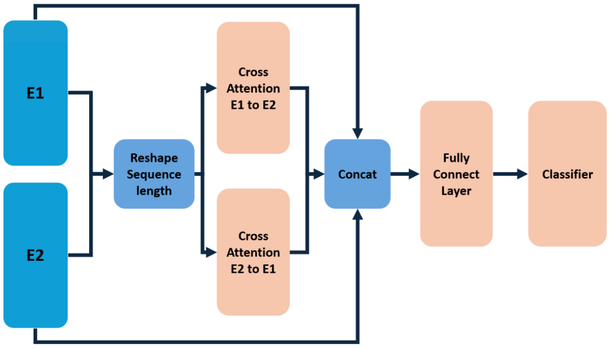


Figure 2. Architecture of the Cross-Attention Model for protein mutation prediction. The model leverages embeddings from two sources, E1 (e.g., AlphaFold 3) and E2 (e.g., ESM2). The embeddings are reshaped to a fixed sequence length and passed through cross-attention layers, where attention is computed from E1 to E2 and vice versa. The outputs are concatenated and processed through a fully connected layer, followed by a classifier to predict the functional impact of mutations.

4. Experiments and Results

4.1. Training Setup

We trained both models using an 80%-20% train-validation split. The training was conducted over 1,500 epochs for the Cross-Attention model and 300 epochs for the GAT model, with a batch size of 32, using the Adam optimizer with a learning rate of 0.001. To prevent overfitting, we employed regularization techniques such as dropout with a rate of 0.5 and early stopping based on validation loss. Training was performed on an NVIDIA GTX 1080 Ti GPU.

4.2. Results

The GAT achieved a validation accuracy of **84%**, marginally outperforming the Cross-Attention model's **83.5%**. Additionally, the GAT demonstrated significant improvements over other recent algorithms in the field. Table 1 shows the comparative performance across multiple datasets, highlighting the robustness of our approach. We compare our method against a diverse set of zero-shot baselines, including the EVmutation model (Hopf et al., 2017), DeepSequence (Riesselman et al., 2018), EVE (Frazer et al., 2021), the CARP suite (Yang et al., 2024), and ESM-IF1 (Hsu et al., 2022).

Our algorithm demonstrates clear advantages, particularly on the **ARGR_ECOLI** dataset, where it significantly outperforms others with an accuracy of **91.8%**. This result underscores the effectiveness of integrating sequence and structure embeddings using a GAT framework.

As shown in Figure 4, the Cross-Attention model demonstrates stable training dynamics.

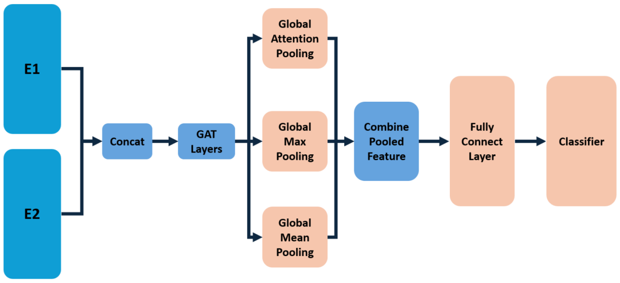
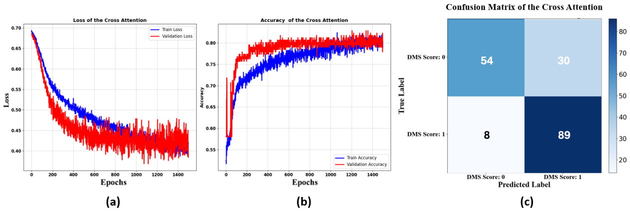


Figure 3. Architecture of the Graph Attention Network (GAT) Model for protein mutation prediction. The model starts by concatenating embeddings from two sources, E1 (e.g., AlphaFold 3) and E2 (e.g., ESM2). These combined embeddings are passed through multiple GAT layers to model relationships in graph-structured data. The outputs are processed using three pooling mechanisms: Global Attention Pooling, Global Max Pooling, and Global Mean Pooling. The pooled features are combined into a single representation, which is passed through a fully connected layer and a classifier to predict the functional impact of protein mutations.



(a) (b) (c)

Figure 4. Performance evaluation of the Cross-Attention Model. (a) Training and validation loss curves over 1,500 epochs, showing a steady decrease in training loss and slight fluctuations in validation loss, indicating stable learning without overfitting. (b) Training and validation accuracy curves, highlighting consistent improvement with convergence around 83% validation accuracy. (c) Confusion matrix on the validation set with an F1-score of 0.82.

Figure 5 illustrates that the GAT model not only achieves higher accuracy but also demonstrates better generalization, as indicated by the confusion matrix.

4.3. Embedding Visualization

We visualized the embeddings at different stages to understand how the models learn to separate classes.

As shown in Figure 6, AlphaFold embeddings cluster by dataset, while ESM2 embeddings partially separate by class. The GAT embeddings demonstrate clear class separation after processing, indicating the model's ability to learn discriminative features.

Table 1. Comparison between our algorithm and recent algorithms on mutation evaluation.

Dataset	EVmutation	EVE	CARP-640M	ESM-If1	DeepSeq	ESM2-650M	ESM2-3B	Ours
LISMN	60	45	60	43	45	43	43	**75**
ARGR_ECOLI	58	54	42	83	55	66	65	**92**
BBC1_YEAST	48	55	47	43	55	54	51	**83**
All	55	54	50	56	52	54	53	**83**

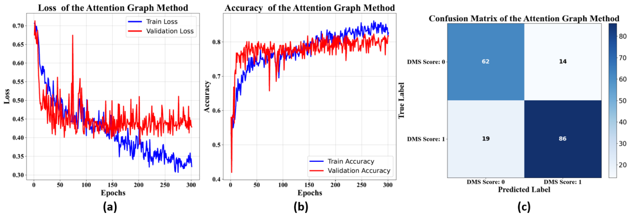


Figure 5. Performance evaluation of the GAT Model. (a) Training and validation loss curves over 300 epochs, showing a significant reduction in training loss and stabilization of validation loss, indicating effective learning. (b) Training and validation accuracy curves, highlighting consistent improvement with convergence close to 84% validation accuracy. (c) Confusion matrix on the validation set with an F1-score of 0.84.

5. Discussion

The superior performance of the GAT model suggests that it more effectively captures the intricate relationships between sequence and structural information. By utilizing graph-based attention mechanisms, the GAT can model nonlinear interactions and dependencies that may be overlooked by the Cross-Attention model. This aligns with our objective of developing a framework that leverages multimodal embeddings to improve mutation impact predictions.

Our results indicate that even with a relatively small dataset, integrating sequence and structure embeddings can significantly enhance predictive performance. The GAT model's robustness across multiple metrics underscores its potential for application in protein engineering and related fields.

6. Conclusion

In this study, we developed a multimodal framework integrating sequence and structure embeddings to predict the impact of protein mutations. Our GAT model outperformed recent methods, achieving a validation accuracy of 84%, and demonstrated the effectiveness of combining embeddings from ESM2 and AlphaFold 3. Future work could explore integrating additional data modalities, such as protein-protein interactions, or applying the framework to larger and more diverse datasets to enhance generalizability.

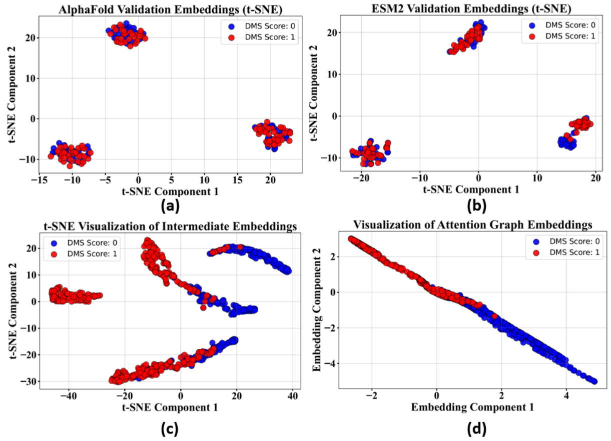


Figure 6. Model embedding visualizations at different stages: (a) t-SNE of AlphaFold validation embeddings (64 dimensions), illustrating clustering based on DMS Score. (b) t-SNE of ESM2 validation embeddings (1280 dimensions), highlighting the class separation. (c) t-SNE of intermediate embeddings after four GAT layers, illustrating feature evolution over layers. (d) Scatter plot of final graph embeddings (2D) after the fully connected layer, showing distinct separation of DMS Score classes.

7. Failed Experiences

We explored contrastive learning (Khosla et al., 2020) to enhance protein graph embeddings but found the classifier effective without it, likely due to dataset size. Our approach included assuming positive samples from one dataset, negatives from others, and implementing a multi-class contrastive framework.

Specific Contributions

Reza Marzban: Conceived the main idea of the study, prepared embeddings using AlphaFold 3, and implemented the GAT model.
Alireza Kheirandish: Obtained 1280-dimensional embeddings using the ESM2 model, implemented the Cross-Attention model, and ran the baseline models on the dataset.
Both Authors: Collaborated on writing the paper.
Supervision: This work was conducted under the supervision of Professor Amirali Aghazadeh.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, pp. 1–3, 2024.

Akdel, M., Pires, D. E., Pardo, E. P., Jänes, J., Zalevsky, A. O., Mészáros, B., Bryant, P., Good, L. L., Laskowski, R. A., Pozzati, G., et al. A structural biology community assessment of alphafold2 applications. *Nature Structural & Molecular Biology*, 29(11):1056–1067, 2022.

Anstrom, D. M., Colip, L., Moshofsky, B., Hatcher, E., and Remington, S. J. Systematic replacement of lysine with glutamine and alanine in escherichia coli malate synthase g: effect on crystallization. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 61(12):1069–1074, 2005.

Cheng, P., Mao, C., Tang, J., Yang, S., Cheng, Y., Wang, W., Gu, Q., Han, W., Chen, H., Li, S., et al. Zero-shot prediction of mutation effects with multimodal deep representation learning guides protein engineering. *Cell Research*, 34(9):630–647, 2024.

Cocco, S., Posani, L., and Monasson, R. Functional effects of mutations in proteins can be predicted and interpreted by guided selection of sequence covariation information. *Proceedings of the National Academy of Sciences*, 121 (26):e2312335121, 2024.

Frazer, J., Notin, P., Dias, M., Gomez, A., Min, J. K., Brock, K., Gal, Y., and Marks, D. S. Disease variant prediction with deep generative models of evolutionary data. *Nature*, 599(7883):91–95, 2021.

Hopf, T. A., Ingraham, J. B., Poelwijk, F. J., Schärfe, C. P., Springer, M., Sander, C., and Marks, D. S. Mutation effects predicted from sequence co-variation. *Nature biotechnology*, 35(2):128–135, 2017.

Hou, R., Chang, H., Ma, B., Shan, S., and Chen, X. Cross attention network for few-shot classification. *Advances in neural information processing systems*, 32, 2019.

Hsu, C., Verkuil, R., Liu, J., Lin, Z., Hie, B., Sercu, T., Lerer, A., and Rives, A. Learning inverse folding from millions of predicted structures. In *International conference on machine learning*, pp. 8946–8970. PMLR, 2022.

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., and Krishnan, D. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020.

Notin, P., Kollasch, A., Ritter, D., Van Niekerk, L., Paul, S., Spinner, H., Rollins, N., Shaw, A., Orenbuch, R., Weitzman, R., et al. Proteingym: Large-scale benchmarks for protein fitness prediction and design. *Advances in Neural Information Processing Systems*, 36, 2024.

Rao, R., Meier, J., Sercu, T., Ovchinnikov, S., and Rives, A. Transformer protein language models are unsupervised structure learners. *Biorxiv*, pp. 2020–12, 2020.

Riesselman, A. J., Ingraham, J. B., and Marks, D. S. Deep generative models of genetic variation capture the effects of mutations. *Nature methods*, 15(10):816–822, 2018.

Seuma, M., Lehner, B., and Bolognesi, B. An atlas of amyloid aggregation: the impact of substitutions, insertions, deletions and truncations on amyloid beta fibril nucleation. *Nature Communications*, 13(1):7084, 2022.

Tang, Q.-Y. and Kaneko, K. Long-range correlation in protein dynamics: Confirmation by structural data and normal mode analysis. *PLoS computational biology*, 16 (2):e1007670, 2020.

Tsuboyama, K., Dauparas, J., Chen, J., Laine, E., Mohseni Behbahani, Y., Weinstein, J. J., Mangan, N. M., Ovchinnikov, S., and Rocklin, G. J. Mega-scale experimental analysis of protein folding stability in biology and design. *Nature*, 620(7973):434–444, 2023.

Yang, K. K., Fusi, N., and Lu, A. X. Convolutions are competitive with transformers for protein sequence pretraining. *Cell Systems*, 15(3):286–294, 2024.

Zeng, W., Dou, Y., Pan, L., Xu, L., and Peng, S. Improving prediction performance of general protein language model by domain-adaptive pretraining on dna-binding protein. *Nature Communications*, 15(1):7838, 2024.