Deep Learning-based Prediction of Protein Conformational Ensembles

Dilara ONAL[1], Onur SERÇİNOĞLU[1]

GEBZE TECHNICAL UNIVERSITY

[1] Bioinformatics and System Biology, Gebze Technical University, 41400, Gebze/Kocaeli, Turkey

HIBIT 2025

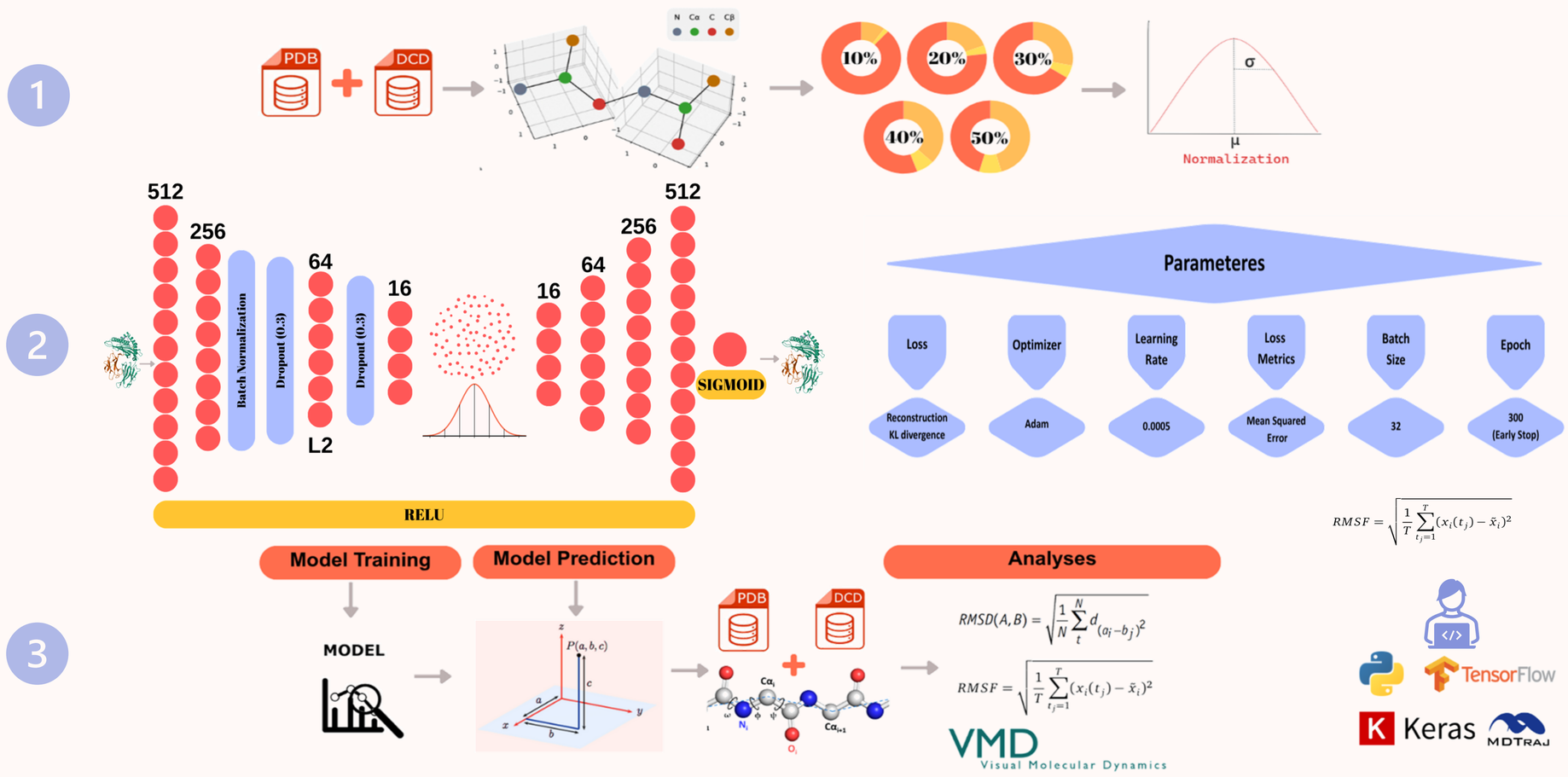
Introduction

Proteins are macromolecules that play a crucial role in organizing biological systems. Understanding their structure and dynamics is critical for disease treatment. Molecular Dynamics (MD) simulations have been used to model the dynamic properties of proteins. These simulations calculate new positions based on the distances and motions of atoms, using phrysical calculations as a reference, and then generate the protein's motions based on these calculations[1]. However, this method also has disadvantages in terms of cost and time.

Recent studies have focused on predicting the dynamic structures of proteins using artificial intelligence[1 2]. The coordinate data obtained from MD simulations have been selected as the data source required for training the deep learning model to reduce simulation time and costs[3]. To solve the running costs and long-term nature of MD simulations, a simulation should generate sufficiently short datasets for model training[4]. This field is relatively new in the literature, and relatively fast-developed methods are available.

In this study, variational autoencoder (VAE) model were used with the obtained MD simulation data, and the results were analyzed to examine how successful they were for dynamic structure estimation.

Methodology



Results

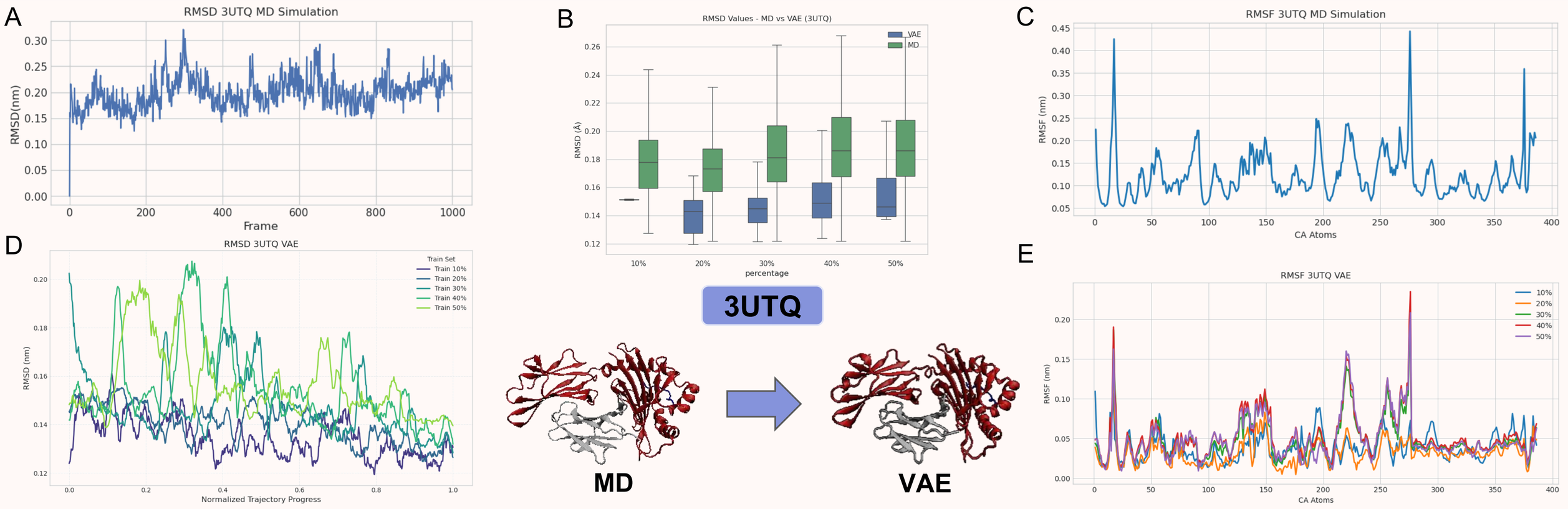


Figure 1. HLA Class I antigens are MHC-encoded glycoproteins expressed on all nucleated cells, presenting intracellular peptides to the immune system. Molecular dynamics (MD) simulation data of the 3UTQ structure were used to perform RMSD and RMSF analyses (A, C). A Variational Autoencoder (VAE) model was trained on MD-derived coordinates with different training ratios, and RMSD and RMSF analyses were also conducted on the model-predicted coordinates (D, E). RMSD values from the MD simulation and VAE predictions were visualized using boxplots (B).

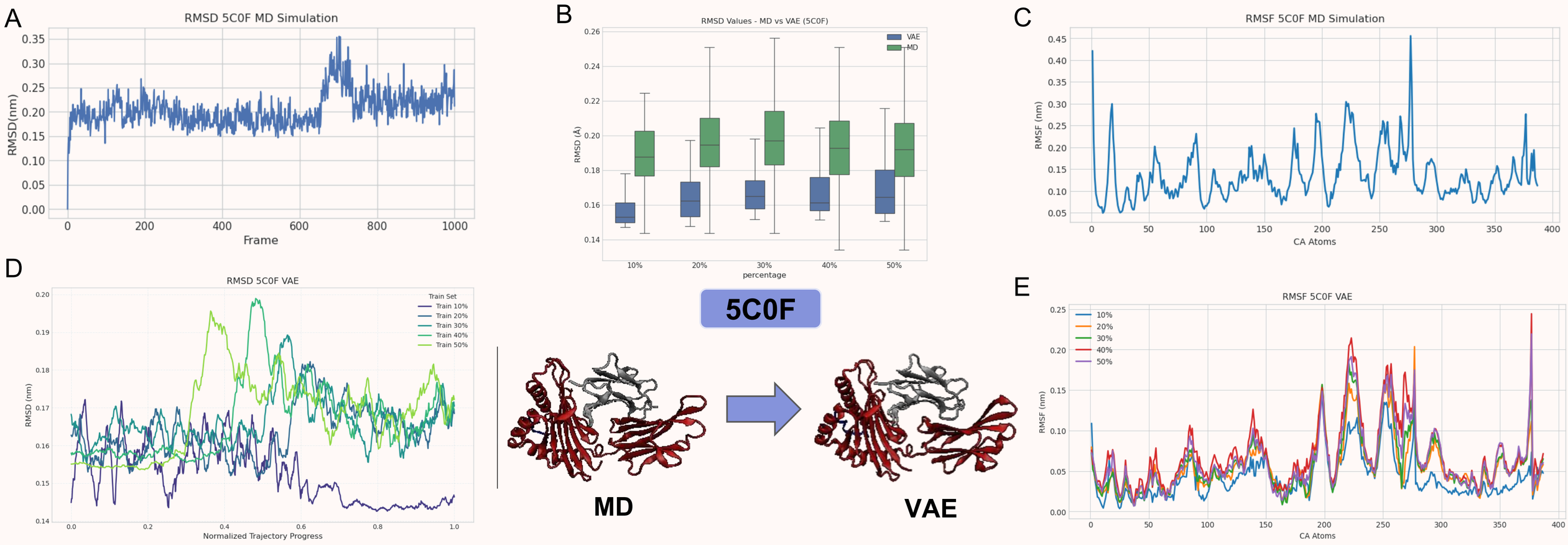


Figure 2.. For the 5C0F structure, molecular dynamics (MD) simulation data were used to perform RMSD and RMSF analyses (A, C). The Variational Autoencoder (VAE) model was trained on MD-derived coordinates with different training ratios, and RMSD and RMSF analyses were performed on the model-predicted coordinates as well (D, E). RMSD values from the MD simulation and VAE predictions were visualized using boxplots (B).

Conclusion

In this study, a Variational Autoencoder (VAE) architecture was developed, and its parameters were optimized based on the obtained results. The model was trained using coordinate data derived from molecular dynamics (MD) simulations of selected proteins. Training datasets were prepared using different proportions of the coordinate data, ranging from 10% to 50%, and the model was trained separately for each case. RMSD and RMSF metrics were employed to evaluate how effectively the model captured conformational changes and atomic fluctuations of the proteins. The results indicate that the most optimal predictions were achieved when 30% and 40% of the data were used for training.

References

1. Badar, M. S., Shamsi, S., Ahmed, J., & Alam, M. A. (2022). Molecular dynamics simulations: concept, methods, and applications. In *Transdisciplinarity* (pp. 131-151). Cham: Springer International Publishing.
2. Degiacomi, M. T. (2019). Coupling molecular dynamics and deep learning to mine protein conformational space. *Structure*, *27*(6), 1034-1040.
3. Gupta, A., Dey, S., Hicks, A., & Zhou, H. X. (2022). Artificial intelligence guided conformational mining of intrinsically disordered proteins. *Communications biology*, *5*(1), 610.
4. Zhu, J. J., Zhang, N. J., Wei, T., & Chen, H. F. (2023). Enhancing conformational sampling for intrinsically disordered and ordered proteins by Variational autoencoder. *International Journal of Molecular Sciences*, *24*(8), 6896.

Contact

Name Surname: Dilara ONAL

e-mail:d.onal2018@gtu.edu.tr