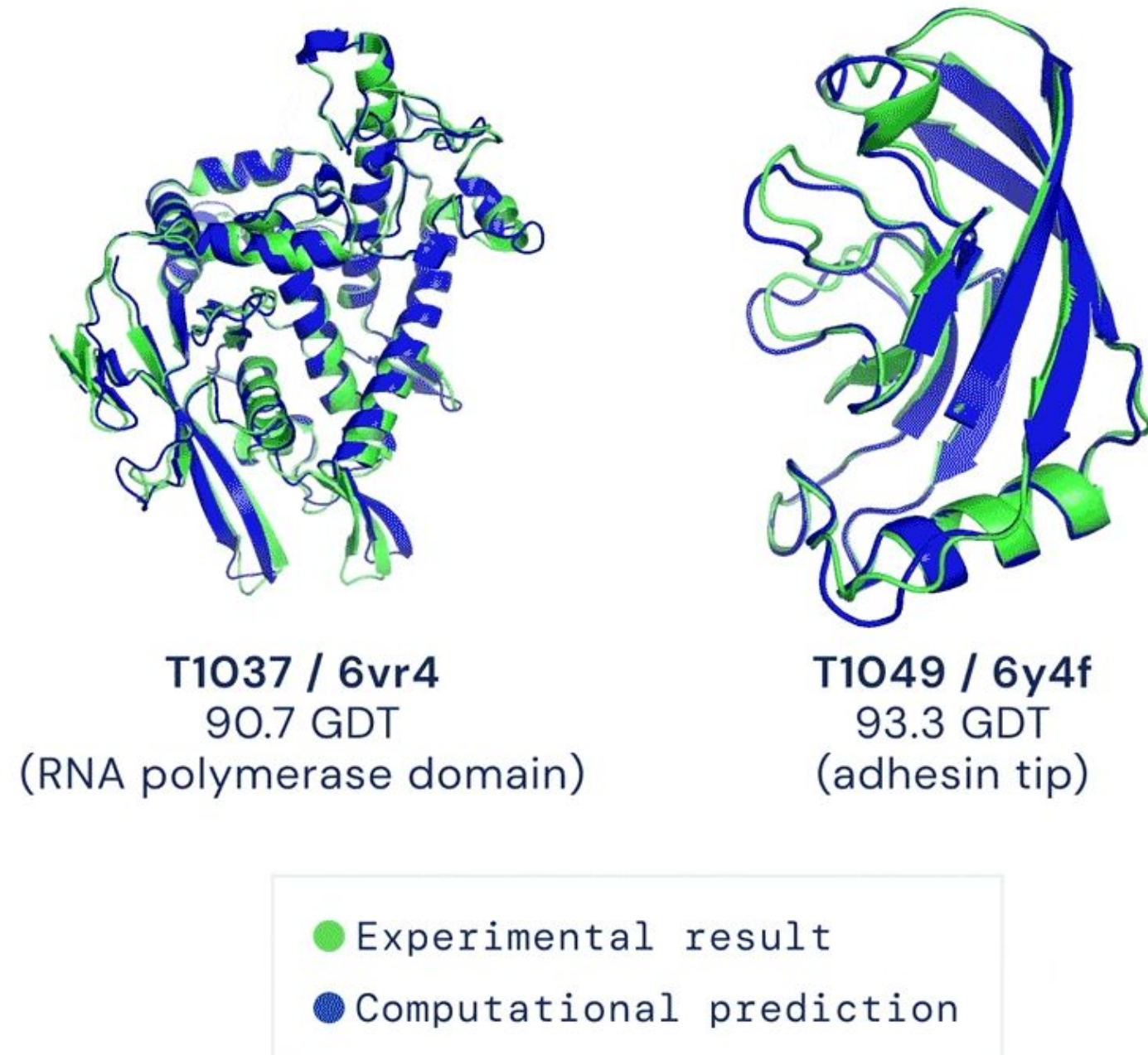


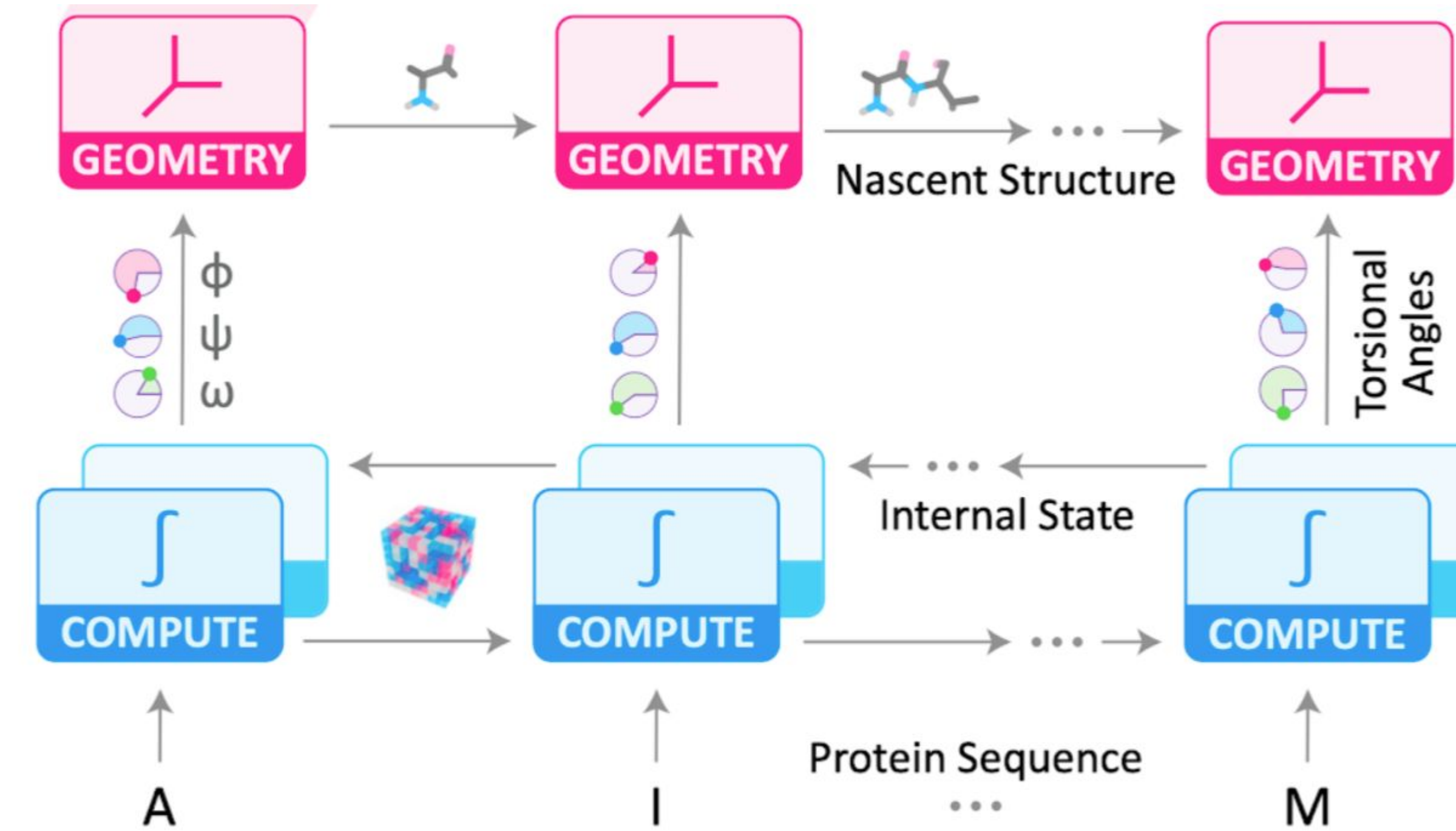
Background



Each protein has a unique 3D shape that determines how it works and what it does. Knowledge of its structure can provide:

- More precise work on drug development
- Useful in future pandemic response efforts

Protein Structure Formation

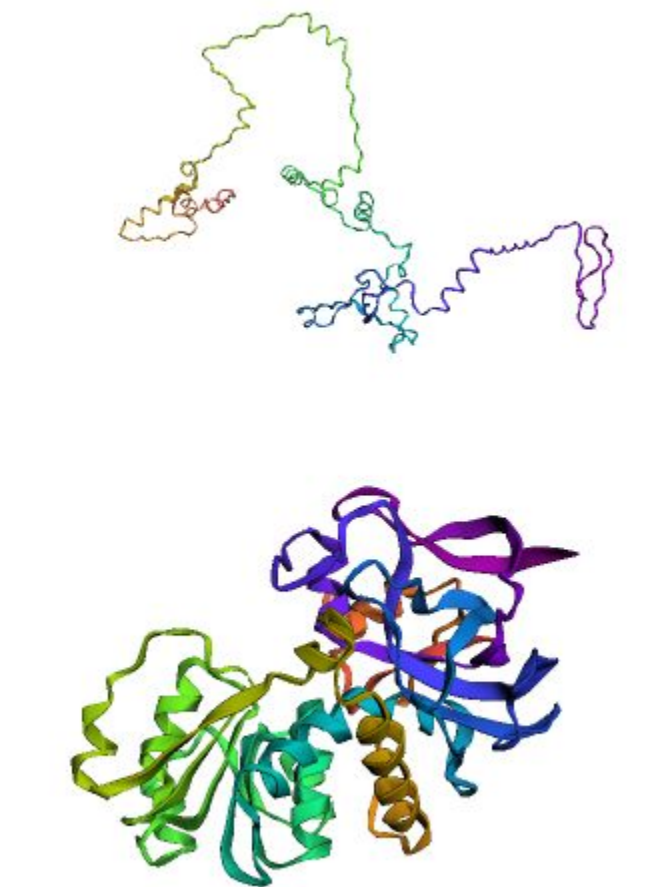
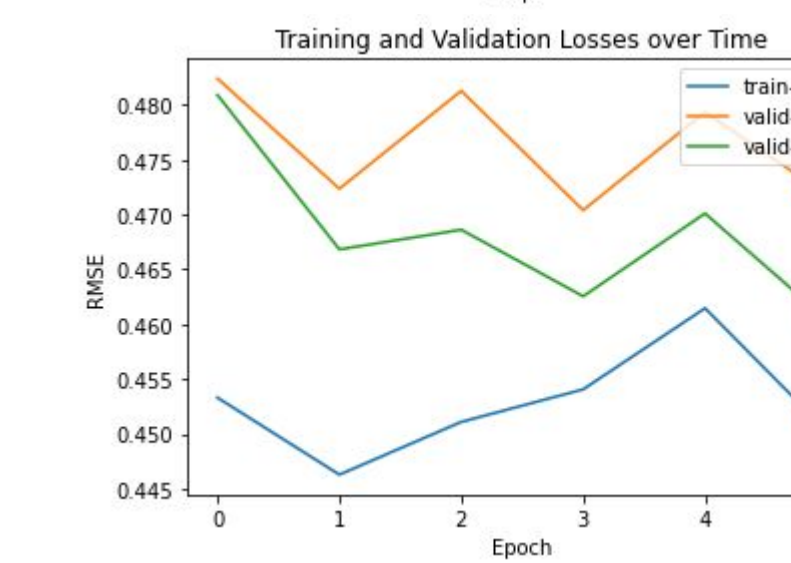


A. Amino acid sequence concatenation
M. Position Specific Scoring Matrix
I. Sidechain information content

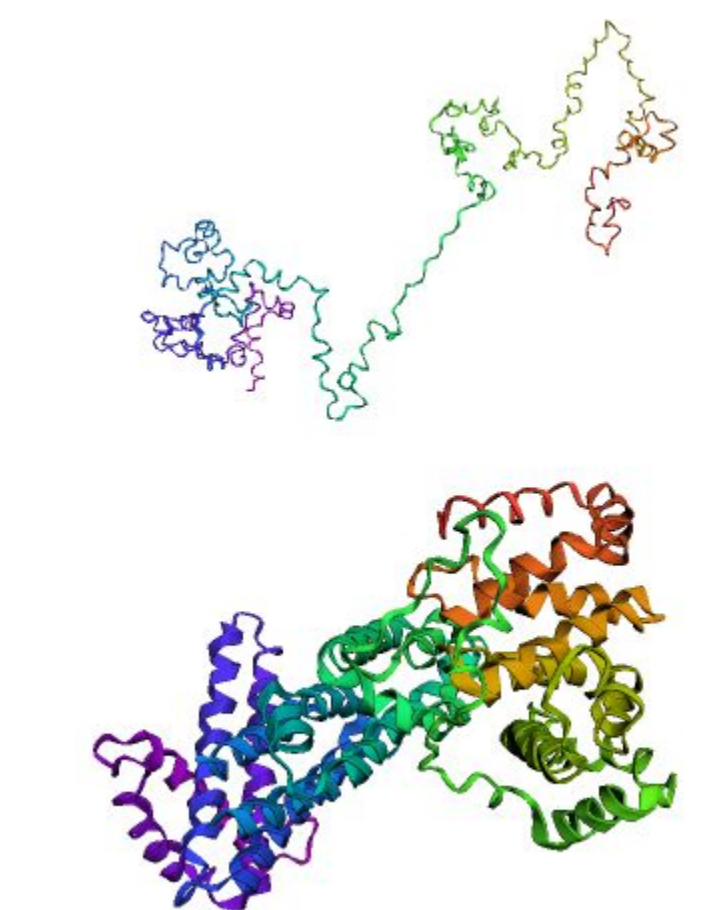
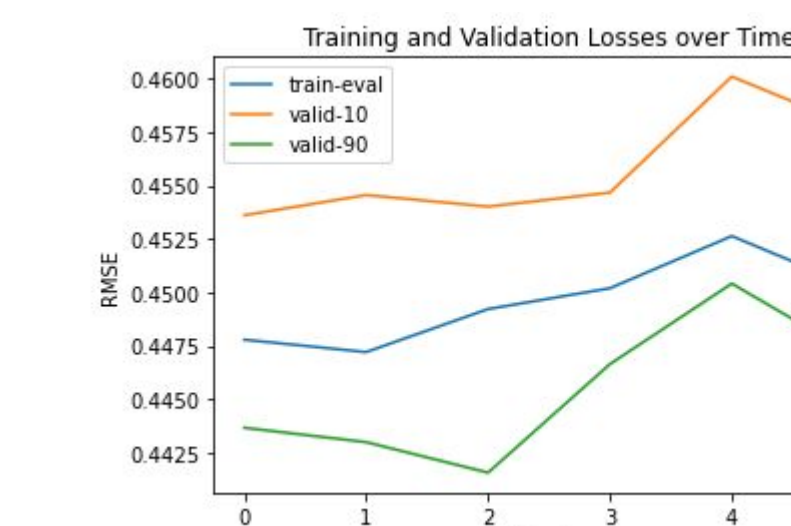
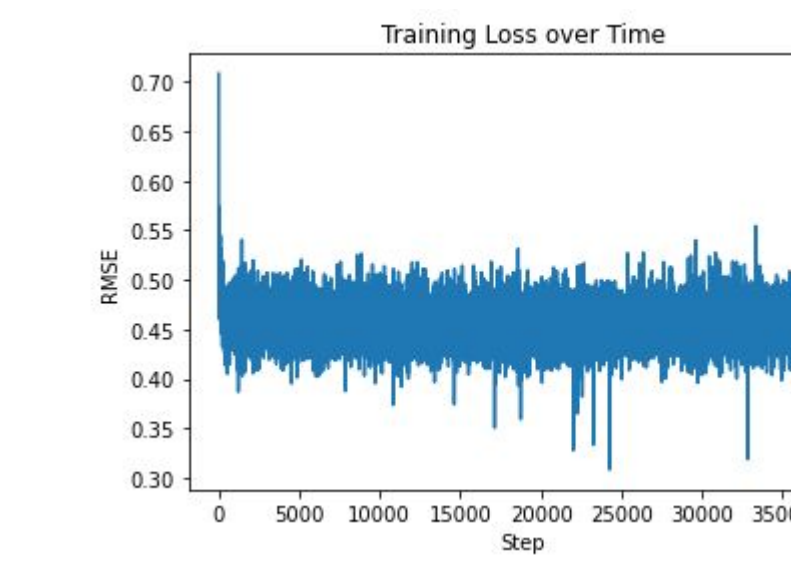
Conclusion and Results



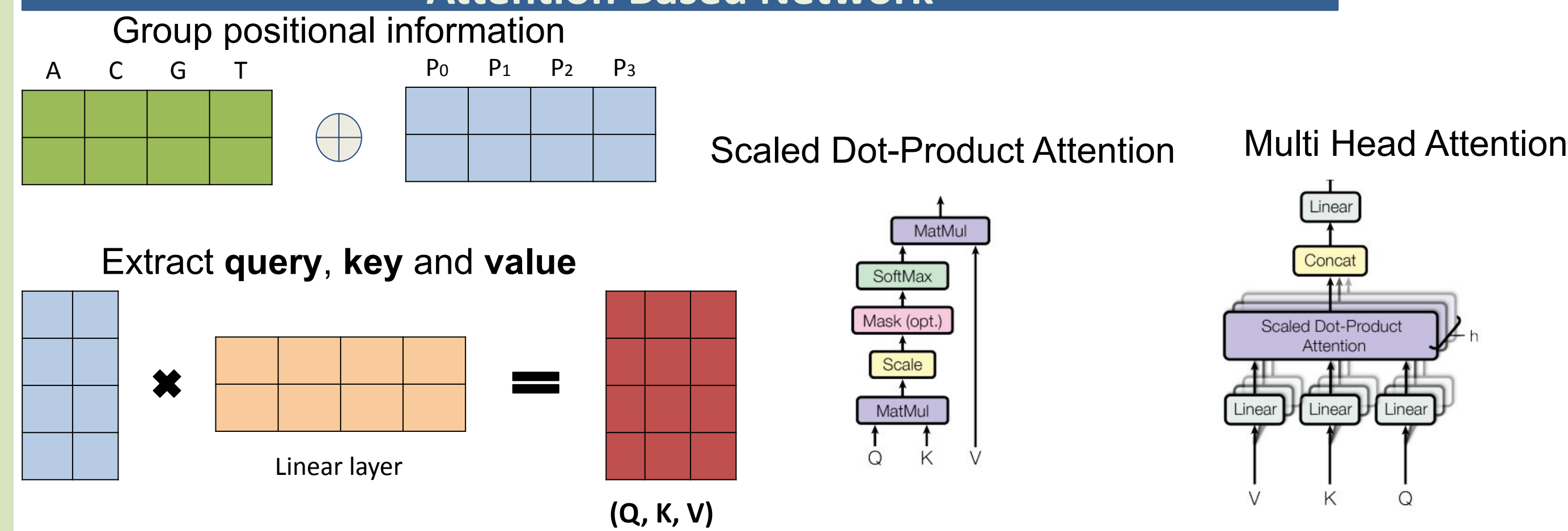
Complete training with sequences, PSSMs and Information content



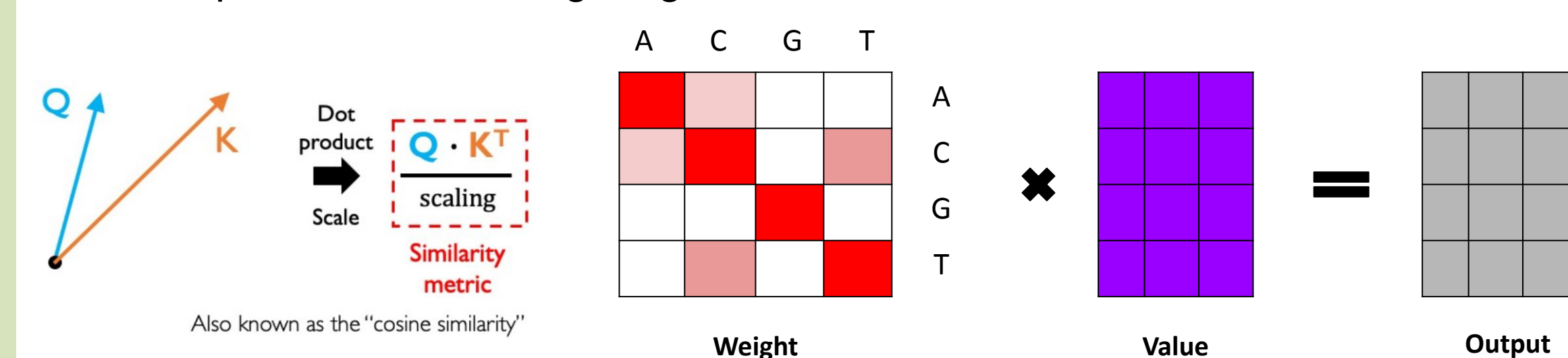
Sequence only training



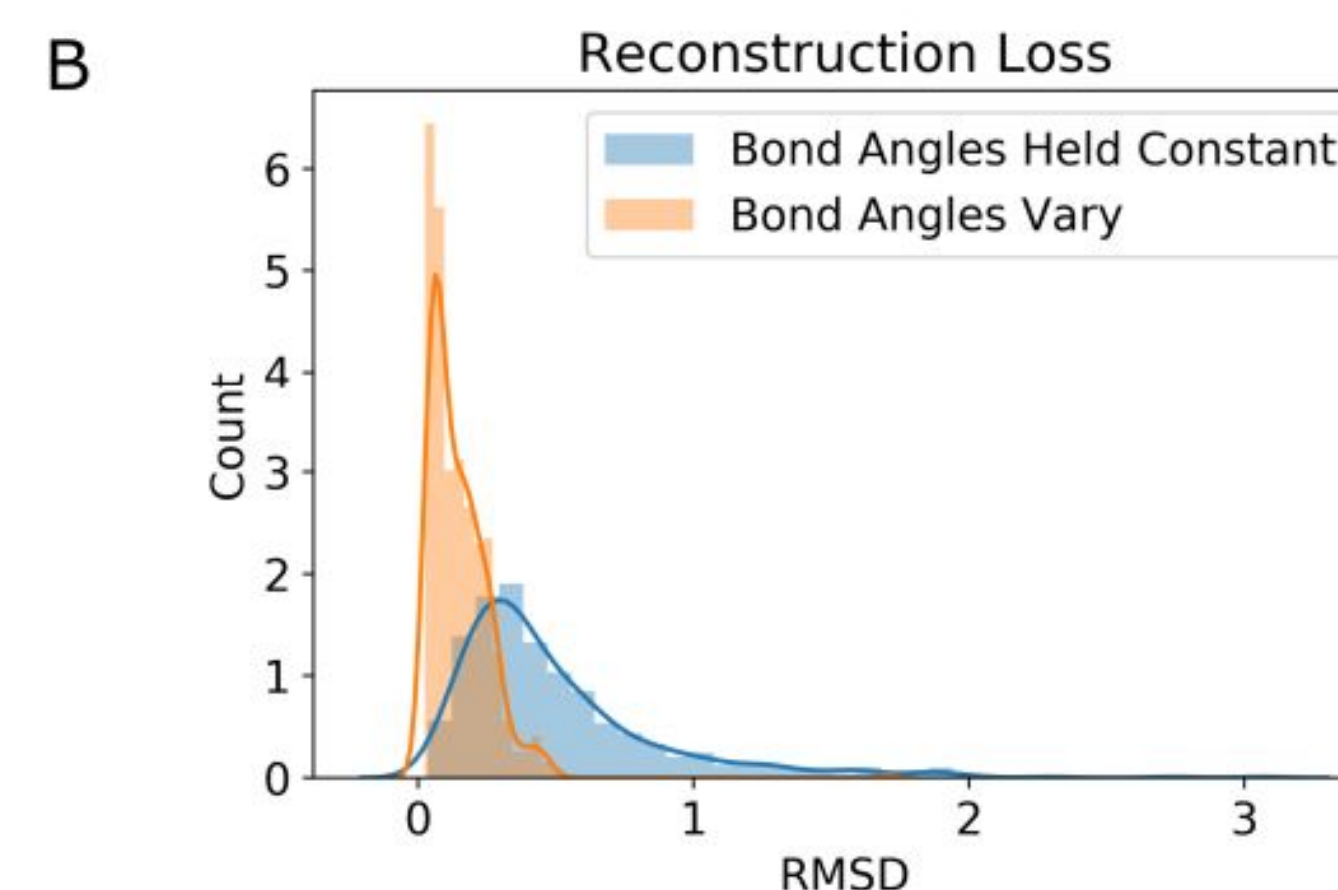
Attention Based Network



Compute Attention Weighting



- A.** Residue with 3 category of angles:
- the canonical backbone torsional angles (Φ , Ψ , and Ω) in green
 - rotatable sidechain-specific torsional angles (X_{1-3}) in purple
 - 3-atom backbone bond angles in orange-pink



B. Change in reconstruction error by considering or not the backbone angles as fixed

Future Work

There are still techniques not implemented that have potential to significantly improve the efficiency of current models, like multimer templates or self-distillation of multimer predictions and assure atomic precision.

Contact

Mateus Antonio Franceschina
BscIT, Concordia University of Edmonton
Email: mafrance@student.concordia.ab.ca

References

- King, Jonathan Edward, and David Ryan Koes. "SidechainNet: An all-atom protein structure dataset for machine learning." *Proteins: Structure, Function, and Bioinformatics* 89.11 (2021): 1489-1496.
- AlQuraishi, Mohammed. "ProteinNet: a standardized data set for machine learning of protein structure." *BMC bioinformatics* 20.1 (2019): 1-10.
- Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021).
- Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).