Mission Statement

My goal in creating AMPForge2 is to engage in frontier research at the intersection of deep learning and bioinformatics. This project will allow me to advance my journey as an independent researcher while refining the models I have previously developed. I hope that AMPForge2 will not only strengthen my research capabilities but also serve as a meaningful exploration of deep learning architectures that can support future antimicrobial peptide (AMP) research at UBC.

Model Improvements

Although I have many ideas for improving upon my previous attempt at this project, the most significant change will be in the model architecture used for prediction. In my original work, I developed a variational autoencoder (VAE) that transformed one-hot encoded amino acids into novel peptide sequences. While this was an excellent exercise in building a generative AI system, it did not reflect the transformative advances introduced by transformer models and their self-attention mechanisms. Moving forward, I plan to rigorously study modern transformer architectures used in leading research labs worldwide in order to design the most relevant system for generating novel AMPs.

Another limitation of my earlier model was the size of the training dataset. In this iteration, I plan to expand the number of sequences from 10,114 to roughly 50,000, while dedicating significantly more effort to sourcing and curating data from multiple AMP databases. This increase in both the quantity and quality of training data will help produce more relevant peptide sequences and reduce the likelihood of overfitting.

Long-Term Considerations

Throughout the lifespan of this project, I aim to maintain a stronger system for version control and more effectively communicate my research, insights, and changes. This time, I plan to document my progress in the form of a blog and structure the project in a way that is both readable and reproducible for students and faculty.

Next Steps

Over the coming weeks, I will begin Phase 1 of the project, which will serve as the foundation. This phase will focus primarily on research and planning, providing the framework needed to design and evaluate future models. My first step will be to gather and study relevant papers from UBC and other leading research groups to refine my understanding of transformer models and their application in generating novel AMP sequences.