Report on Protein-Ligand Modeling for the ASAP Discovery Antiviral Ligand Poses 2025 Competition

1. Introduction

The ASAP Discovery Antiviral Ligand Poses 2025 Competition challenges participants to predict ligand binding poses for MERS-CoV Mpro and SARS-CoV-2 Mpro using structural data. This report outlines the development and application of a specialized protein-ligand modeling approach, inspired by AlphaFold3's architecture, with a significant emphasis on data processing and ligand diversity analysis.

2. Methodology

2.1 Data Analysis and Ligand Diversity

An initial analysis of the provided training and test datasets focused on ligand diversity to identify common substructures. This analysis revealed that the quinoline scaffold was prevalent in the majority of samples. Quinoline is recognized as a privileged substructure in antimicrobial drugs due to its versatile binding capabilities.

2.2 Data Filtering and Preparation

To enhance model training:

- Binding Pocket Focus: Samples outside the main binding pocket were excluded to concentrate on relevant interactions.

- Ligand Similarity Filtering: Ligands dissimilar to those in the test set were removed, resulting in a refined training set of 266 samples.

- Validation Set Selection: Despite the reduced dataset size, a diverse validation set was curated, emphasizing samples most similar to the test set to ensure robust model evaluation.

2.3 Model Development

The model development was inspired by AlphaFold3's architecture, which incorporates a diffusion-based generative model operating on raw atom coordinates. This approach allows the model to refine "fuzzy" initial predictions into well-defined structures. By integrating the quinoline scaffold as a focal point, the model was tuned to predict ligand poses that align with expected binding conformations.

3. Results

The specialized model demonstrated improved accuracy in predicting ligand poses, particularly for quinoline-containing ligands. This targeted approach ensured that the quinoline scaffold was consistently placed in the anticipated binding pose, enhancing the reliability of the predictions.

4. Conclusion

By focusing on ligand substructure analysis and leveraging advanced modeling techniques inspired by AlphaFold3, the developed model effectively predicts ligand binding poses for MERS-CoV Mpro and SARS-CoV-2 Mpro. This methodology underscores the importance of integrating domain-specific knowledge with cutting-edge computational architectures in structure-based drug discovery.