Proposal of Reinforcement Learning Project: **Integrating ESM3 Structural Representation with Self-Play Reinforcement Learning**

Group member: Yifan Qin, Yan Zeng, Haizhao Dai

Abstract

This project proposes to combine ESM3's structure tokenization technology with EvoPlay's Monte Carlo Tree Search (MCTS) self-play reinforcement learning method to develop a novel protein backbone design approach. ESM3 compresses protein local structures into discrete tokens through Vector Quantized Variational Autoencoder (VQ-VAE), enabling effective representation and reconstruction of protein structures; meanwhile, EvoPlay utilizes policy-value networks to guide MCTS in efficiently exploring protein sequence space. By integrating these two methods, we will develop a protein design system capable of direct optimization at the structural level, achieving protein backbone generation from first principles, and addressing limitations of traditional sequence or fragment assembly methods. This project will provide important technical support for protein design.

Related Works

Protein design has been a core challenge in biotechnology. Traditional methods primarily based on sequence mutations or fragment assembly often struggle to effectively explore protein structure space. Recent advances for Reinforcement learning in protein design struggles in using helical and loop fragments to design part of a symmetrical assembly, where the design space is less extensive. Other advances in artificial intelligence, particularly AlphaFold2 (AF2) and ESM series language models, have brought revolutionary breakthroughs to protein structure prediction and design.

ESM3 introduced structure tokenization technology, compressing complex 3D structural information into discrete tokens via VQ-VAE, with each token containing geometric information about a residue and its surrounding environment. This approach not only allows high-precision structure reconstruction (backbone RMSD < 1Å) but also provides new avenues for structure generation.

Simultaneously, EvoPlay adopted self-play reinforcement learning methods derived from AlphaZero, efficiently exploring protein sequence space through MCTS guided by policy-value networks. This method has demonstrated advantages in multiple protein engineering tasks, including full-length protein optimization, peptide binder design, and site-directed combinatorial mutations.

Integrating these two methods will create a design system that optimizes directly at the structural level, potentially overcoming limitations of existing methods and achieving more effective protein backbone design.

Method

1. Definition of Structure Token Action Space
    - Utilize ESM3's VQ-VAE encoder to map structures to 4,096 possible structure tokens
    - Define an action space based on structure token changes, enabling MCTS to directly modify protein backbones
    - Develop structure validity checking mechanisms to ensure generated structure token sequences can be decoded into valid protein backbones
2. Policy-Value Network Architecture Design
    - Design a policy-value network applicable to structure tokens, evaluating structural states and guiding search
    - Employ deep neural networks to process structure token sequences, predicting action probabilities and state values
    - Train policy networks through self-play, optimizing the balance between structure exploration and exploitation
3. Structure Evaluation and Reward Function Design
    - Use ESM3's structure decoder to convert structure tokens to 3D coordinates
    - Integrate AlphaFold2's scoring components (such as pLDDT and PAE) or more physically-based scoring function to evaluate the quality of generated structures
    - Design comprehensive reward functions considering structural quality, stability, and packing density
4. Hierarchical MCTS Implementation
    - Implement hierarchical MCTS algorithms, with higher-level decisions responsible for structure token optimization and lower-level decisions for amino acid sequence optimization
    - Develop effective search strategies balancing structure space exploration with local optimization
    - Implement parallel MCTS to improve computational efficiency and accelerate the design process

Objective

1. Reproduce the results in EvoPlay on protein sequences and the results of ESM3 structure tokenization on protein backbone generation.
2. Develop a structure token-based protein backbone generation framework integrating ESM3's structure tokenization with EvoPlay's MCTS method
3. Establish a policy-value network suitable for structure token optimization and design effective reward functions to evaluate the quality of generated structures

Potential Risks

1. Large structure token space leading to low search efficiency
2. High computational resource requirements

Reference

1. Thomas Hayes et al. ,Simulating 500 million years of evolution with a language model.Science387,850-858(2025).DOI:10.1126/science.ads0018
2. Wang, Y., Tang, H., Huang, L. et al. Self-play reinforcement learning guides protein engineering. Nat Mach Intell 5, 845–860 (2023). https://doi.org/10.1038/s42256-023-00691-9
3. Isaac D. Lutz et al. ,Top-down design of protein architectures with reinforcement learning.Science380,266-273(2023).DOI:10.1126/science.adf6591