

I. Multimodal Alignment

(General Protein, PDB)

Caption:

Chain A is a **single-chain protein** that tends to be **helical** with H: 20-30, E: 20-30, and C: 50-60. The **longest run** for H is 9-15. It contains more than five **helix segments** and more than five **strand segments**.

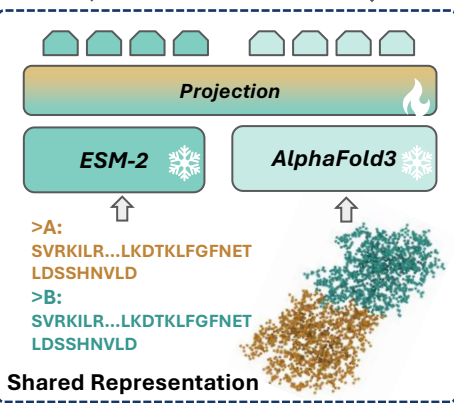


Understanding Experts



Understanding Tokenizer

Describe **secondary structure information, including major SS, SS fraction, longest run, longest count** for protein 1szz with chain A and chain B.

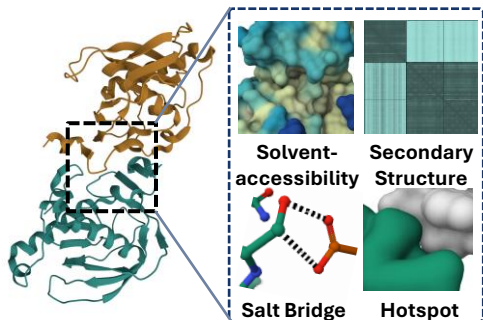


Shared Representation

>A:
SVRKILR...LKDTKLFQFNEL
LDSSHNVLVD
>B:
SVRKILR...LKDTKLFQFNEL
LDSSHNVLVD

II. Mid-Training

(General Protein, PDB)



Understanding Experts



Understanding Tokenizer

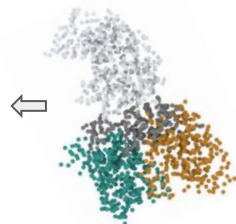
M0: What are the **solvent-accessibility** bins ...
M1: What is the **secondary structure labels** between ...
M2: For chain A and B, what is the **salt-bridge count** ...
M3: which residues are the **top-5 hotspots** under ...

III. Joint Training

(Antibody, SabDab)



Self-conditioning



Generation Expert



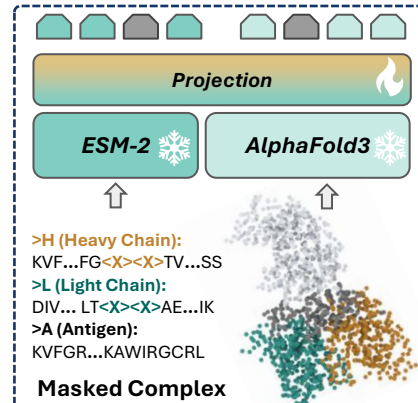
Conditioning Trunk



Understanding Experts



Understanding Tokenizer



You are **redesigning masked CDR** of an antibody to improve binding to the antigen. Design points (antigen hotspots): **[A,114], [A,146], [A,119]**