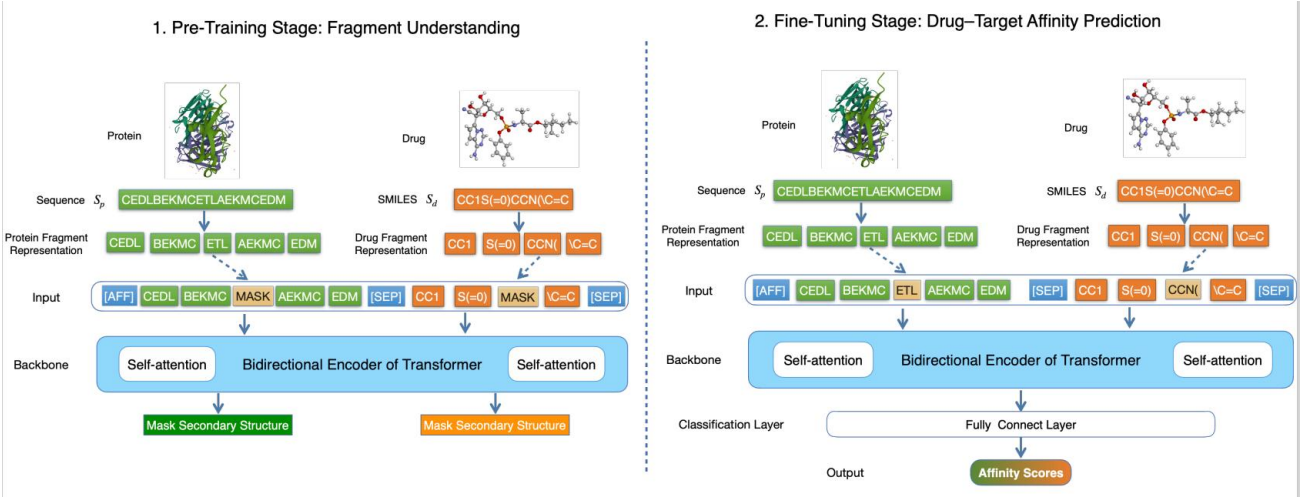
FragDPI: A Novel Drug–Protein Interaction Prediction Model Based on Fragment Understanding and Unified Coding

Zhihui YANG, Juan LIU, Xuekai ZHU, Feng YANG, Qiang ZHANG, Hayat Ali Shah

Frontiers of Computer Science, DOI: 10.1007/s11704-022-2163-9

Problems & Ideas

- Problems of conventional stereo matching approaches:
  - They extract features from protein and drug sequences separately, they can not learn the features characterizing the drug-protein interactions.
  - They encode the protein (drug) sequence usually based on the assumption that each amino acid (atom) has the same contribution to the binding, ignoring different impacts of different amino acids (atoms) on the binding.
- Ideas:
  - We propose to predict drug-protein binding with a novel unified coding strategy so that the interaction information can be learnt.
  - we propose to encode proteins and drugs by using conserved fragments instead of single amino acids or atoms.



Main Contributions

- Contributions:
  - A unified coding strategy was introduced that tokenizes the sequence of a protein by identifying conserved fragments in the sequence;
  - We encode drugs and proteins into a unified vector to more precisely describe the details of drug-protein interactions and simplify the encoding process;
  - We build the model through a novel two-step training strategy, the pre-training stage and the fine-tuning stage.