LINKER: Learning Interactions Between Functional Groups and Residues With Chemical Knowledge-Enhanced Reasoning and Explainability

Phuc Pham[1, *] Viet Thanh Duy Nguyen[1] Truong-Son Hy[1, †]

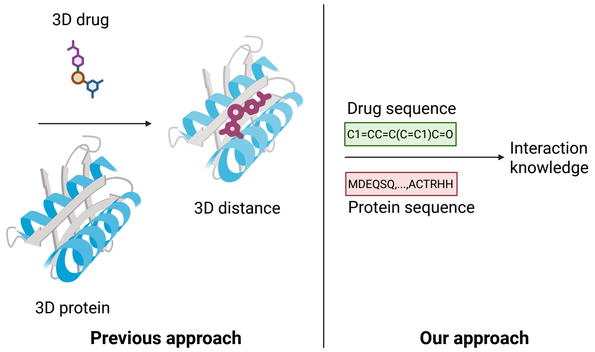
[1]Department of Computer Science, The University of Alabama at Birmingham, USA

[*]This work is done during the author's research internship at the University of Alabama at Birmingham under the guidance of Dr. Truong-Son Hy

[†]Corresponding author: thy@uab.edu

HySonLab
AI + Sciences

Introduction

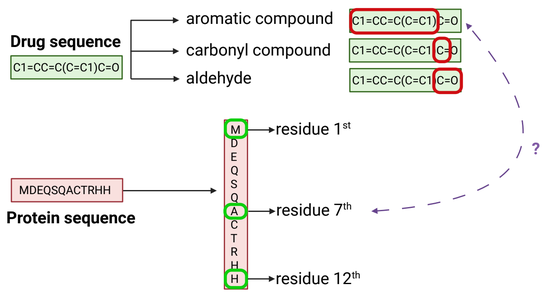


- Proteins & drugs interact like locks and keys → knowing the exact interaction is critical for drug design.
- Current approach needs 3D structures (rare) or only checks closeness (limited insight).
- Our approach: sequence-based model that predicts meaningful drug-protein interactions directly from sequence data.

Contribution

- First sequence-based framework for predicting biologically meaningful types (H-bonds, π-stacking, salt bridges).
- Sequence-level inference: no 3D structure, scalable & interpretable.

Problem Statement

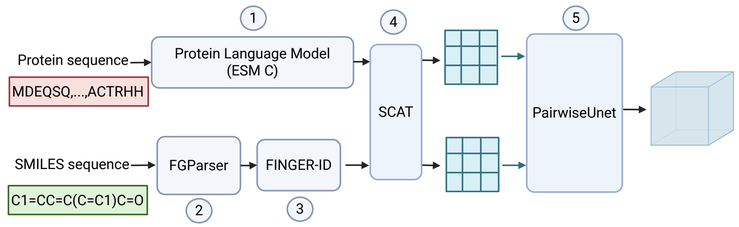


- Does the aromatic group at that position interact with residue 7 of the protein?
- What type of interaction is it?

- Input:
  - Protein sequence: **T** (with R residues).
  - Drug sequence: **D** (with F functional groups).
- Output: probabilities over 7 interaction types, including H-bond, hydrophobic, π-stacking, π-cation, salt bridge, water bridge, halogen bond.

$$f_{\text{LINKER}} : (\mathbf{T}, \mathbf{D}) \rightarrow \mathbf{P} \in [0,1]^{R \times F \times 7}$$

Architecture



(1) Protein Language Model (ESM C) encodes protein sequences; (2) FGParser extracts functional groups from SMILES; (3) FINGER-ID builds context-aware functional group embeddings; (4) SCAT: self & cross attention for residue–group integration; (5) PairwiseUNet predicts interaction probability maps.

Residue Interaction Prediction

- LINKER predicts a probability tensor $\mathbf{P} \in [0,1]^{R \times F \times 7}$ (R: residues, F: functional groups, 7: interaction types).
- **Step 1:** Aggregate over functional groups *(from* $\mathbf{P}$*)*

$$\mathbf{U}_{r,k} = \max_{1 \le f \le F} \mathbf{P}_{r,f,k}, \quad 1 \le r \le R,\ 1 \le k \le 7$$

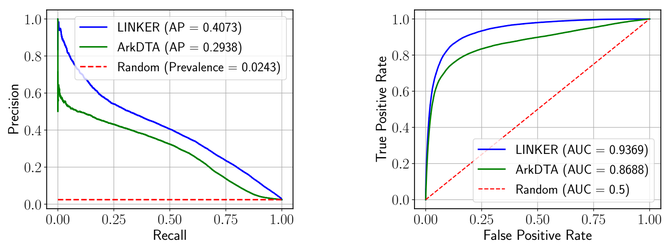
We have residue-interaction matrix $\mathbf{U} \in [0,1]^{R \times 7}$.

- **Step 2:** Aggregate over interaction types *(from* $\mathbf{U}$*)*

$$\mathbf{y}_r = \max_{1 \le k \le 7} \mathbf{U}_{r,k}, \quad 1 \le r \le R$$

Final residue-level vector $\mathbf{y} \in [0,1]^R$ (residue prediction probabilities).

- Enables fair comparison with ArkDTA (residue-level supervision).

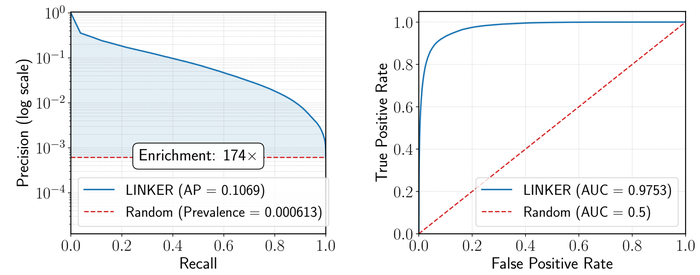


(a) Precision–Recall (b) ROC Curve

Residue-level interaction prediction. LINKER consistently surpasses ArkDTA in both PR *(left)* and ROC *(right)*. The PR analysis emphasizes robustness under class imbalance, while the ROC highlights stronger overall discrimination.

Residue-Functional Group Interaction Prediction

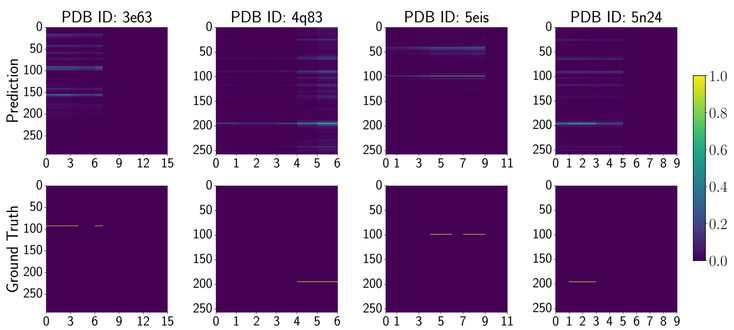
Quantitative Evaluation:



(a) PR curve with enrichment over the prevalence baseline (dashed). (b) ROC curve showing strong discrimination between interacting and non-interacting pairs.

Residue-Functional Group interaction prediction. LINKER delivers markedly higher enrichment at low recall and strong overall discrimination compared to a random baseline.

Quanlitative Evaluation:



Comparison between LINKER predictions and PLIP ground truth for hydrophobic contacts across four protein structures. X-axis: functional group indices from FGParser; Y-axis: residue indices.

Transferability of LINKER Representations to Binding Affinity Prediction

Model	Train	Validation	Test
AutoDock Vina	2.42	2.29	2.56
InteractionGraphNet (IGN)	1.65	2.00	2.16
Random Forest (RF)-Score	0.68	2.14	2.10
DeepDTA	1.41	2.07	2.29
MPRL	**0.48**	**1.47**	1.55
ArkDTA	1.18	**1.47**	1.48
LINKER (Binding Affinity Predictor)	1.38	1.53	**1.47**

Comparison of RMSE on the Leak-Proof PDBBind benchmark for binding affinity prediction.