PLaNet-X 단백질 리간드 결합 부위 및 결합 친화도 예측 모델 개발

소속 정보컴퓨터공학부

분과 A

으쌰으쌰♥

김다현, 박주은, 안수현

지도교수 송길태

연구 배경 및목표



Docking

현재 단백질–리간드 결합 부위 및 친화도 예측은 각광 받고 있음 그러나 기존 연구는 3D 구조 데이터 의존성이라는 한계 존재

단백질 구조 확보 어려움 구조 정보가 없는 단백질적용 불가

활용 가능한 데이터 제한 신약 후보 탐색 속도 저하 실제 신약 개발 단계에서 효율성 감소

• 단백질 서열 + 리간드 구조(2D/SMILES)만으로 결합 부위와 친화도 예측

• 3D 구조 의존성을 줄이고, 빠르고 효율적인 스크리닝을 가능하게 함

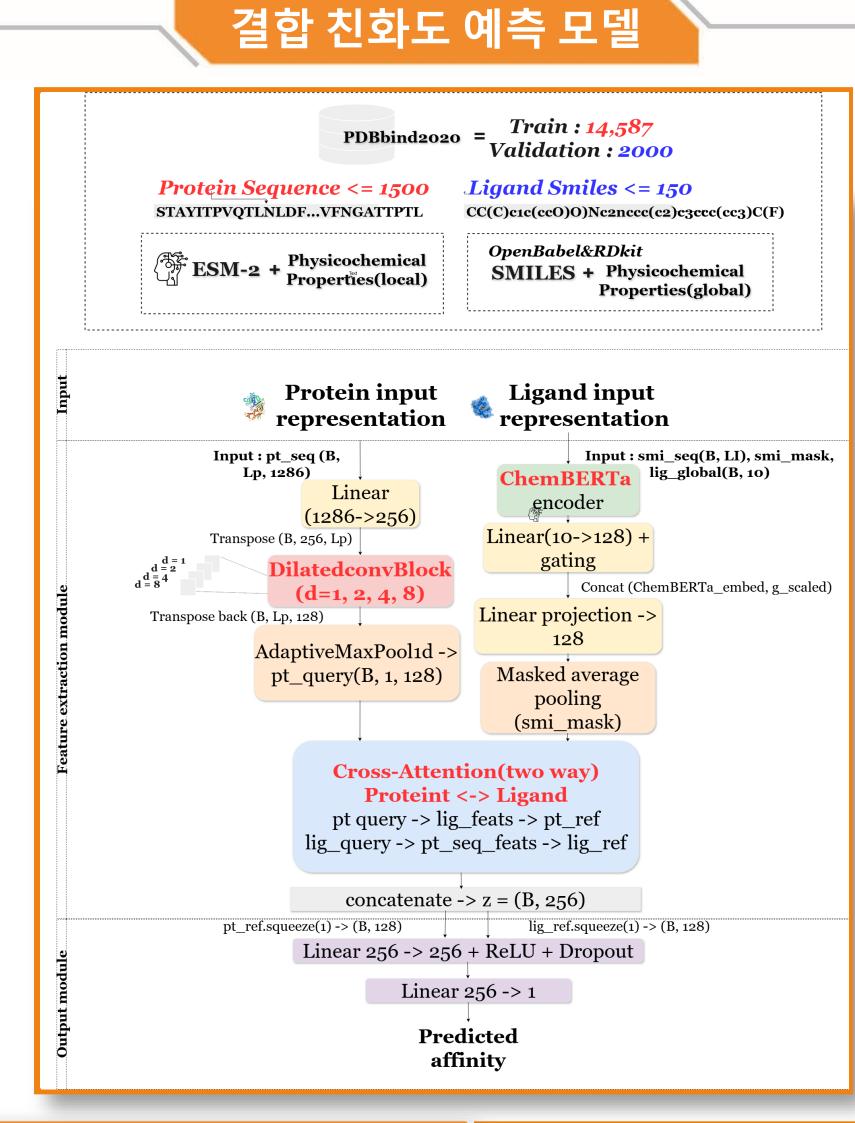
• 신약 개발 초기 단계에서 실질적 활용성을 높임

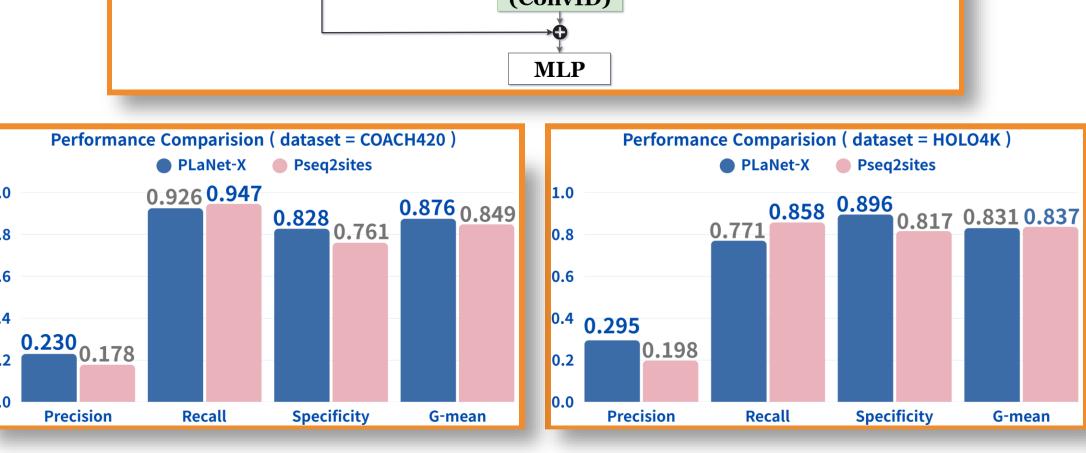
Binding Site Then... How STRONG?

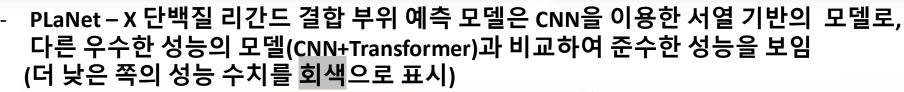
Binding Affinity

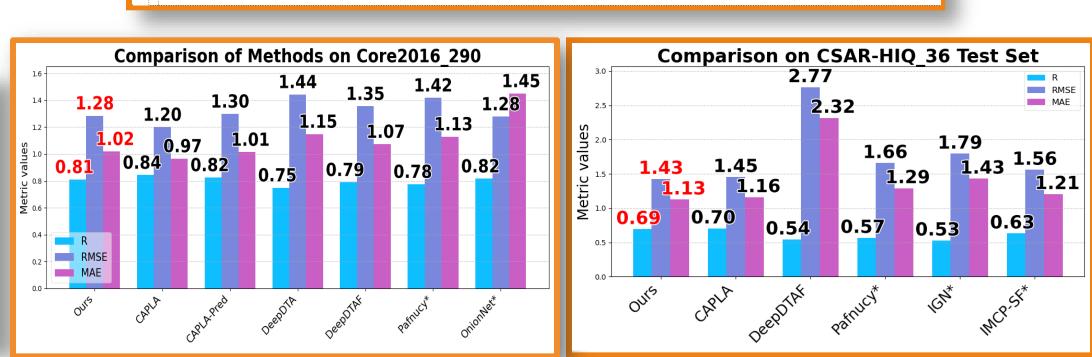
결합 부위 예측 모델

25162 PDBbind =scPDB + **Protein Sequence** STGSATTTPIDSLDDAYITPVQIGTPAQTLNLDF...VFNGATTPTL <=1500 ESM-2 d = 1280**Bottleneck** Conv₁d (k=1, 128)**Bottleneck Bottleneck** Conv₁d Conv₁d (k=5, d=2, 256) (k=3, d=1, 256)**Bottleneck** Bottleneck GAP (global average Conv₁d Conv₁d pooling) (k=3, d=2, 256) (k=5, d=4, 256)**Bottleneck Bottleneck** d = 1280Conv₁d Conv1d Length = 1 (k=3, d=3, 256)(k=5, d=6, 256) **Projection Projection** (Conv₁D) (Conv₁D) **Projection** (Conv₁D) MLP









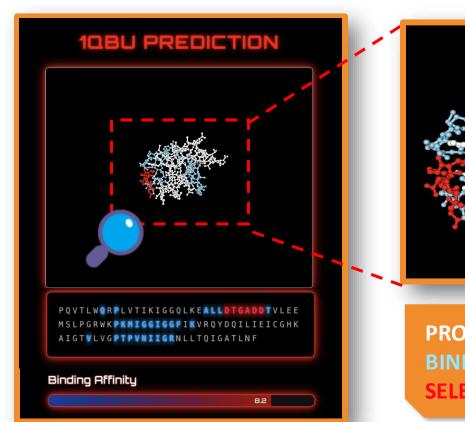
PLaNet – X 단백질 리간드 결합 친화도 예측 모델은 CNN+Transformer를 이용한 서열 기반의 친화도 모델로 타 모델(구조 기반 및 입력 정보 多)들과 비교하여 준수한 성능을 보임

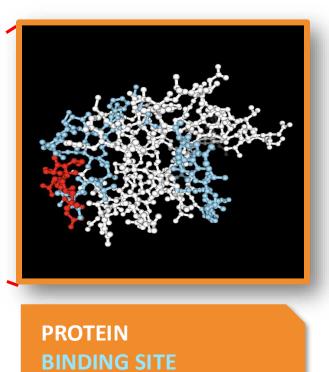
사용자 입력 및 시각화

서비스 구조

aws MY SQL **NVIDIA** Flask Flask

Protein Name f Protein Sequence (FASTA) % Ligand SMILES





SELECTED BINDING SITE