2025년도 연합학습 기반 신약개발 가속화 프로젝트 사업

멀티모달 화합물-전사체-단백질 통합 파운데이션 모델 기반 고정밀 ADMET 예측 기술 개발

주관기관: 숭실대학교

공동기관1: 한국화학연구원
공동기관2: ㈜온코크로스

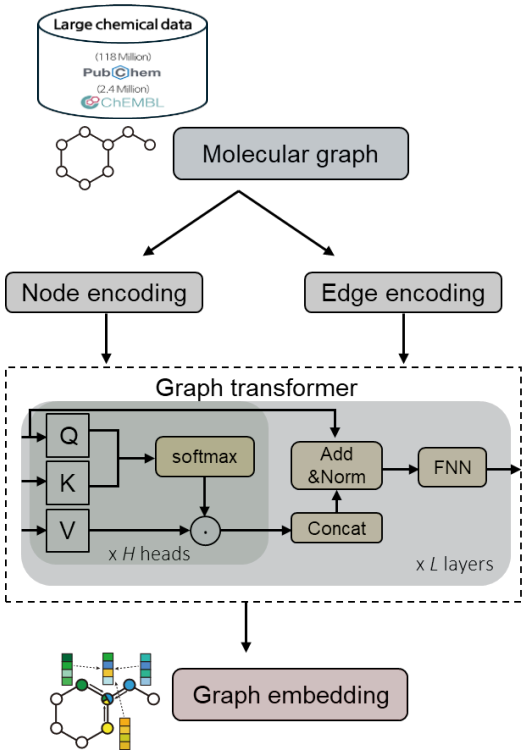
주관책임자
류재용 (숭실대학교)

공동책임자
장우대 (한국화학연구원)

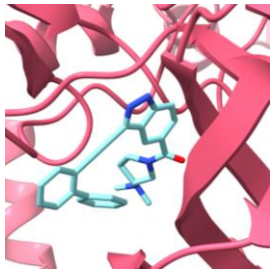
공동책임자
김이랑 (㈜온코크로스)

멀티모달 화합물-전사체-단백질 통합 파운데이션 모델 기반 고정밀 ADMET 예측 기술 개발

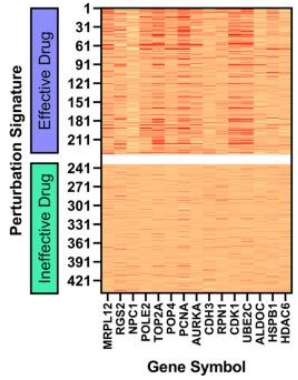
화합물 구조 기반 단일 모달리티 파운데이션 모델의 한계



1. 단백질 상호작용 정보 부재



2. 유전자 발현 정보 부재



멀티 모달리티 기반 고정밀 ADMET 예측 전략

1. 화학 구조 기반 파운데이션 모델

- 화학 구조 기반 파운데이션 모델 구축
- *In vivo* PK 특화 모델 구축

2. 단백질 상호작용 기반 파운데이션 모델

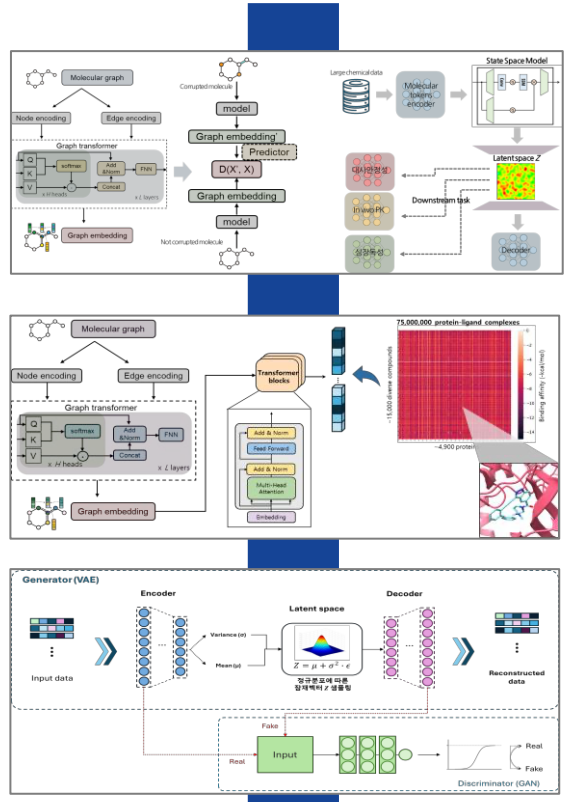
- 단백질-리간드 결합 데이터 분석
- 단백질-리간드 결합 파운데이션 모델 구축

3. 전사체 데이터 기반 파운데이션 모델

- 전사체 파운데이션 모델 구축
- 약물-구조 전사체 기반 파운데이션 모델 구축

4. 선행 연구 결과 기반 모델 고도화

- 선행 ADMET 전문 모델을 활용한 pseudo labelling



멀티모달 파운데이션 모델 기반 ADMET 예측

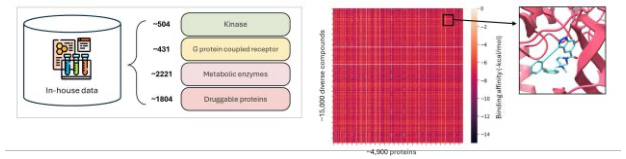
연구개발 목표

▪ 최종 목표 : 멀티모달 화합물-전사체-단백질 통합 파운데이션 모델 기반 고정밀 ADMET 예측 기술 개발

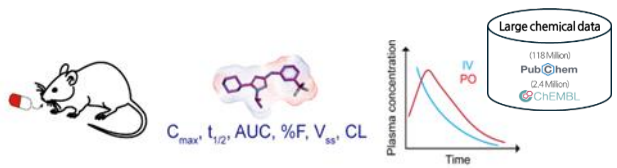
1차년도

연구개발 목표

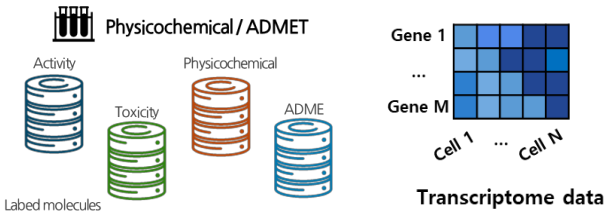
- 멀티모달 화합물-전사체-단백질 통합 파운데이션 모델 구축을 위한 데이터 수집 및 분석



단백질-리간드 상호작용 데이터 분석



In vivo PK 데이터 확보 및 전처리

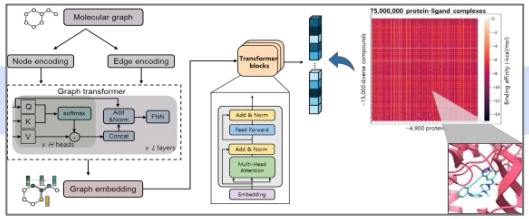


약물 유도 전사체 데이터 확보 및 전처리

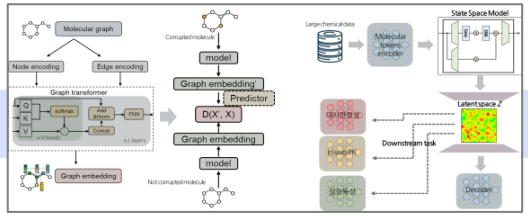
2차년도

연구개발 목표

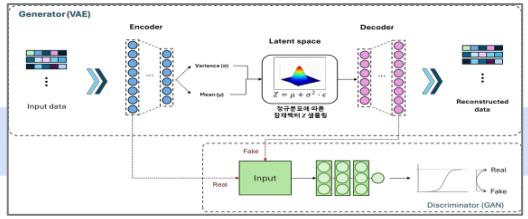
- 멀티모달 화합물-전사체-단백질 통합 파운데이션 모델 구축



단백질 상호작용 데이터 기반 파운데이션 모델 구축



화학 구조 기반 파운데이션 모델 구축



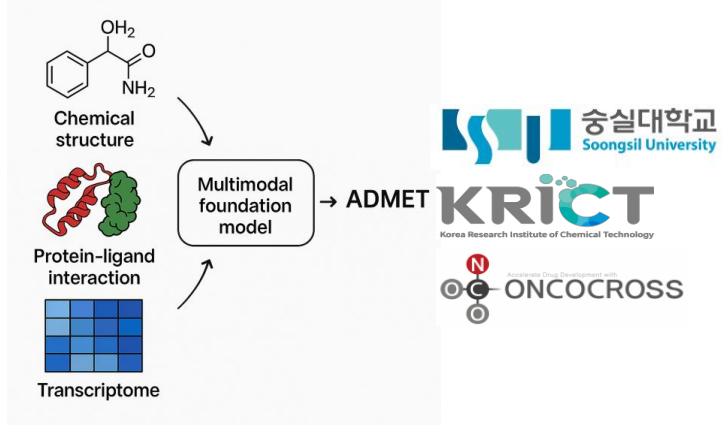
전사체 데이터 기반 파운데이션 모델 구측

3차년도

연구개발 목표

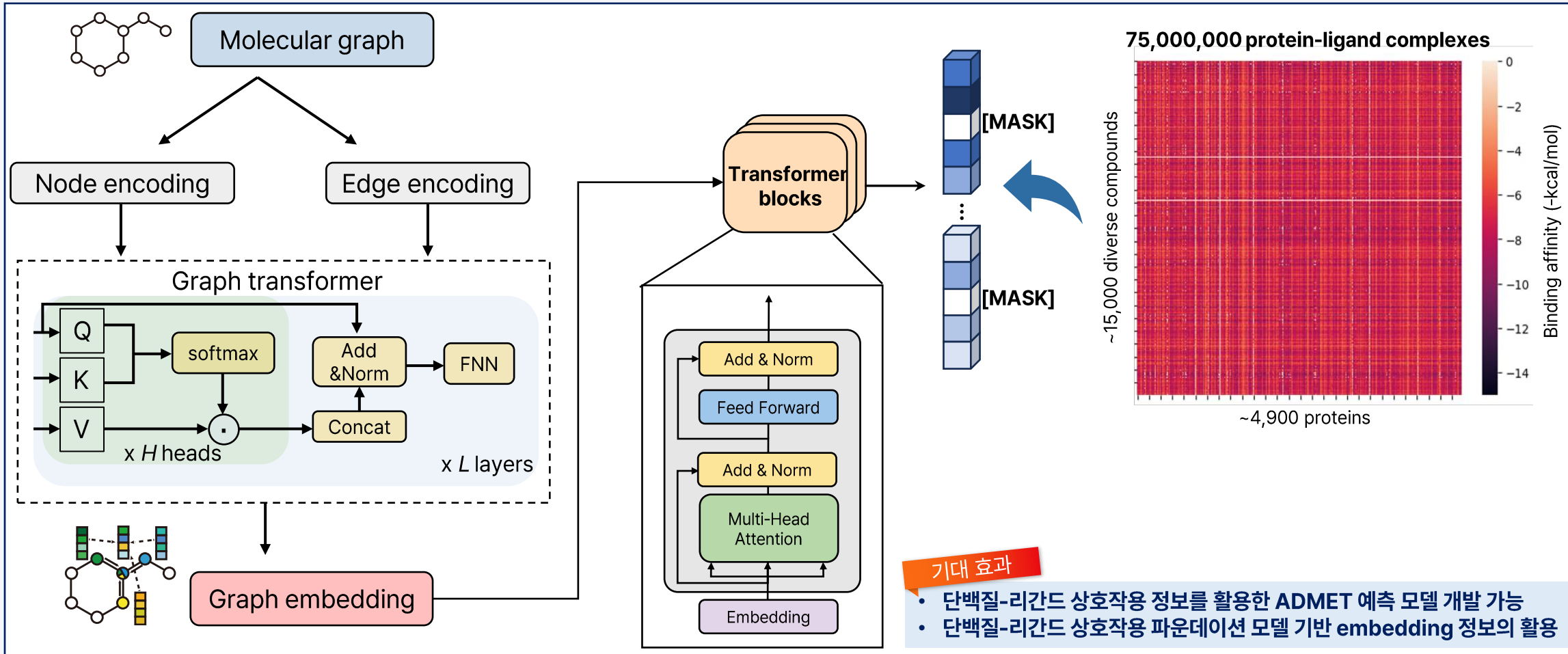
- 멀티모달 화합물-전사체-단백질 통합 파운데이션 모델 최적화

Mixture of experts (MoE)



멀티 모달리티 기반 통합 파운데이션 모델 최적화 및 검증

단백질-리간드 복합체 파운데이션 모델 구축



기대 효과

- 단백질-리간드 상호작용 정보를 활용한 ADMET 예측 모델 개발 가능
- 단백질-리간드 상호작용 파운데이션 모델 기반 embedding 정보의 활용

멀티모달 파운데이션 모델 구축

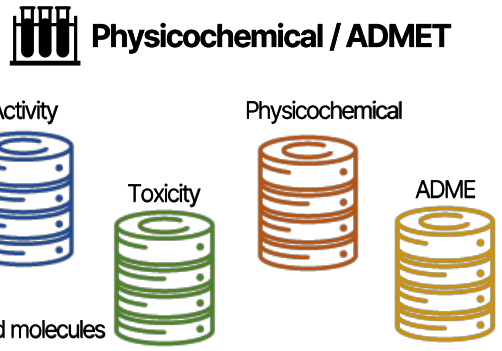
멀티모달 화합물-전사체-단백질 통합 파운데이션 모델 기반 ADMET 예측 모델 구축

선행 연구 결과 기반 모델 고도화

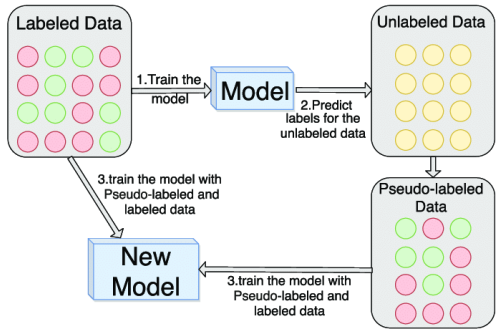
Mixture of experts (MoE)

약물 상호작용 예측 (PNAS, 2018)	심장 독성 예측 (Bioinformatics, 2020)	대사 안정성 예측 (Bioinformatics, 2022)	급성 흡입 독성 예측 (BMC Bioinformatics, 2023)	약물 상호작용 예측 (PNAS, 2023)	혈장 안정성 (CSBJ, 2023)

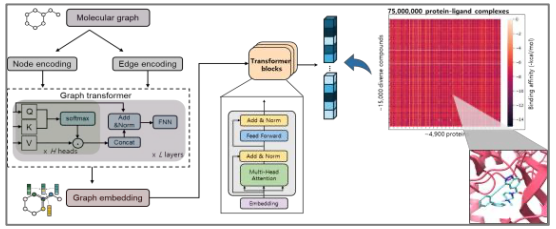
In-house database



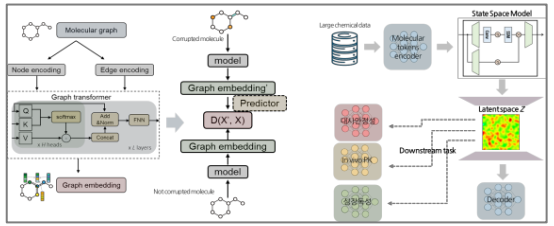
Pseudo-labeling



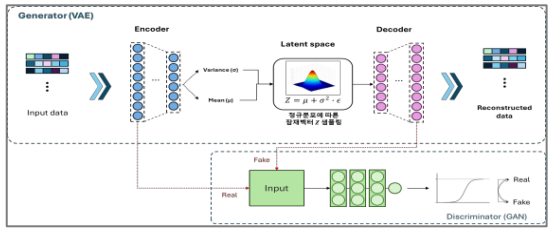
화학 구조 기반 파운데이션 모델

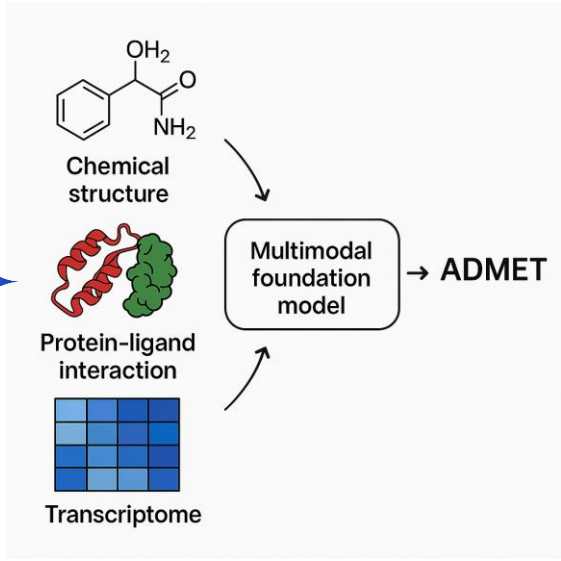


단백질 상호작용 기반 파운데이션 모델



전사체 데이터 기반 파운데이션 모델



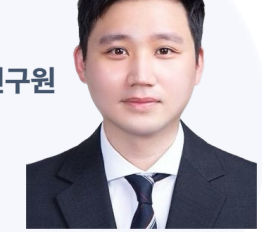


공동연구책임자 연구역량

인공지능 및 컴퓨터 기반 신약개발 분야 전문가

장 우 대

- 한국화학연구원 선임연구원



주요 약력

- 한국화학연구원 신약정보기술연구센터 선임연구원
- KAIST 생명화학공학과 공학박사
- UST 의약화학 및 약리생물학 조교수
- (前) KAIST 생물공정연구센터 연구원

연구 실적 및 주요 성과

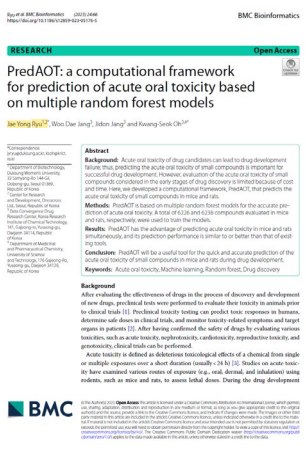
- 인공지능 및 컴퓨터 기반 신약개발 관련 우수 연구성과 보유
- PNAS를 포함한 SCIE급 논문 총 23편 게재 (주저자: 11편), 특허 출원 4건/등록 2건
- 과기부 'AI 활용 혁신신약 발굴 사업' 및 '국가신약개발사업' 과제 수행 중
- 다수 제약사와 AI 활용 신약개발 프로젝트 수행 중 (유한양행, 동아ST-레드엔비아, 타스컴 등)
- 국책 연구소 전략과제 '인공지능 기반 ADMET 예측 플랫폼' 구축 프로젝트 수행 중
- ADMET 모델 구축으로 논문2편 게재, 국내 특허 2건 출원

혈장 안정성 예측 모델 (CSBJ, 2023)

Computational and Structural Biotechnology Journal

PredPS: Attention-based graph neural network for predicting stability of compounds in human plasma

급성 독성 예측 모델 (BMC Bioinformatics, 2023)


BMC Bioinformatics

RESEARCH Open Access

PredAOT: a computational framework for prediction of acute oral toxicity based on multiple random forest models

코로나바이러스 치료제 개발 (PNAS, 2021)

Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay

항암제 개발 (Biomedicines, 2022)

biomedicines

CDI Exerts Anti-Tumor Effects by Blocking the FoxM1-DNA Interaction

코로나바이러스 치료제 (국내 특허 등록, 2023)

특허증 CERTIFICATE OF PATENT

특 허 제 10-2609352 호

출원번호 제 10-2021-0055736 호

출원일 2021년 04월 29일

등록일 2023년 11월 29일

위의 발명은 「특허법」에 따라 특허원부에 등록되었음을 증명합니다.

This is to certify that, in accordance with the Patent Act, a patent for the invention has been registered at the Korean Intellectual Property Office.

2023년 11월 29일

특허청장 COMMISSIONER, KOREAN INTELLECTUAL PROPERTY OFFICE

특허청 Korean Intellectual Property Office

공동연구개발기관 연구수행 능력

한국화학연구원은 ADMET/PK 기본모델 구축, 데이터 전처리 도구 개발 수행

AI 신약개발 전문가와 신약개발/약동력학 전문가로 연구팀 구성

모델 학습용 계산 자원 및 웹서비스

자체적인 AI 모델 구축 장비 확보 및 웹서비스 경험 보유

<로컬 계산 자원 확보>

<ADMET 웹 서비스 운영 경험>

System	CPU	GPU	Application
SuperServer 740GP-TNRT	Intel(R) Xeon(R) Gold 6150 CPU @ 2.70GHz 36 core	Tesla V100 (4EA)	데이터 전처리 도구 개발
SuperServer SYS-740GP-TNRT	Intel(R) Xeon(R) Gold 6348 CPU @ 2.60GHz 56 core	RTX A6000 (4EA)	ADMET/PK 기본모델 구축
SuperServer 740GP-TNRT	Intel(R) Xeon(R) Gold 6242R CPU @ 3.10GHz 80 core	Tesla V100 (4EA)	
SuperServer SYS-521GE-TNRT	Intel(R) Xeon(R) Gold 6526 CPU @ 2.80GHz 32 core	RTX PRO6000 (3EA), 7월 도입 예정	

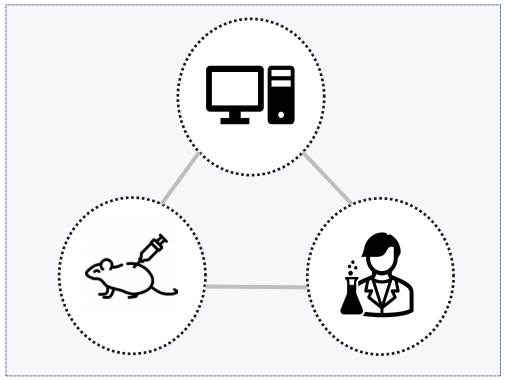
연구진 구성

AI 모델 및 약동력학 도메인 전문가로 구성

장우대 박사 (공동연구과제 책임자)

- ✓ 한국화학연구원 선임연구원
- ✓ KAIST 생명화학공학 박사

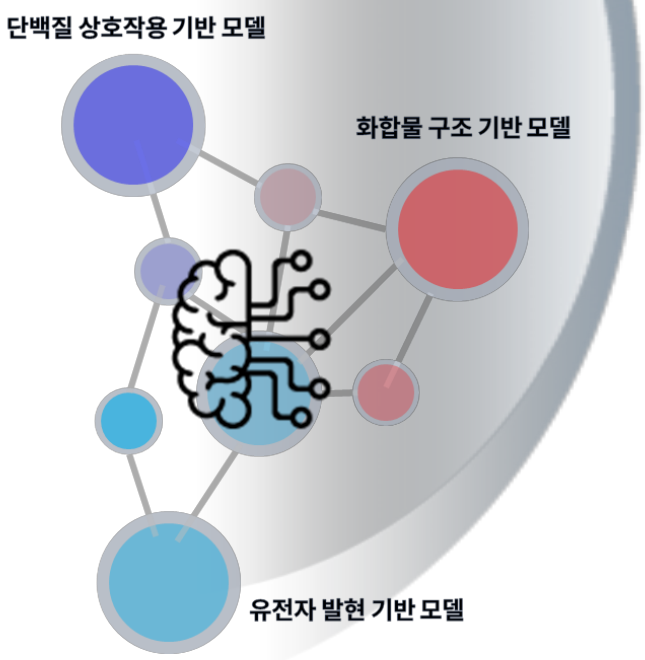
AI 신약개발 전문가



- AI 모델 개발 및 신약개발/약동력학 전문가로 연구팀 구성
- 신약개발/약동력학 분야에서 10년 이상의 연구 경험을 가진 참여 연구원 (박사급 1명, 석사급 1명)
- 신약개발 도메인 지식을 활용한 데이터 품질 관리, 모델 검증 및 최적화

마일스톤 목표 및 정성/정량 목표

멀티모달 화합물-전사체-단백질 통합 파운데이션 모델 기반 고정밀 ADMET 예측 기술 개발



마일스톤

1차년도: 파운데이션 모델용 데이터셋 준비 및 태스크 맞춤형 전처리 도구 개발

- 파운데이션 모델용 데이터셋 확보
- 태스크에 따른 맞춤형 전처리 도구 개발

2차년도: 멀티 모달리티 기반 파운데이션 모델 개발

- 멀티 모달리티 기반 파운데이션 모델 데이터 확보
- ADMET 태스크 모델 개발에 활용 가능한 멀티 모달리티 기반 파운데이션 모델 구축

3차년도: FAM 기본 모델 개발 및 성능 검증

- AI 기반 ADMET 예측 모델(FAM) 개발
- 멀티 태스크 전략 기반 성능 고도화
- 연합학습 기반 모델 고도화

정성/정량 목표

성과지표 및 목표치

- ADMET 및 PK 예측 기본 모델 개발 3건 이상
- ADMET 태스크 성능 개선 모델 확보 2건 이상
- SCI(E) 학술지 등록 건수: JCR 10% 이내 포함 3편
- 특허 출원건수: 3건
- 데이터 전처리 도구를 포함한 SW 등록 2건 및 인증 2건 이상
- 추가 공개 데이터 공급 보고서 1건 이상 제작
- FAM 솔루션 사용자 매뉴얼 1건 이상 제작
- FAM 솔루션 활용 선도물질 최적화 성공 사례 2건 이상 달성

성능지표 및 측정방법

- 분류 모델: AUROC 기본모델 대비 5% 이상 증가
- 회귀 모델: RMSE 기본모델 대비 5% 이상 감소