

# 신약 개발 AI 모델의 정확도 향상을 위한 고품질 데이터 제공 및 최적화

2024. 8. 20



## 신약개발에 집중하는 R&D 중심 제약기업

- 신약 후보물질 **기술수출 및 공동 연구** 경험을 통해 글로벌 수준의 R&D 역량을 확보
- 비만/대사, 항암, 희귀질환 분야를 중심으로 **총 26개의 신약 파이프라인** 보유
- 한국 및 중국에 **6개의 R&D센터**가 유기적으로 협업 진행

### 서울 연구센터 (본사)

- 서울시 송파구
- 임상연구 및 개발, 허가등록, 라이선싱
- R&D 인력: 134



### 북경한미 R&D센터

- 중국 북경
- 이중항체 신약 연구
- R&D 인력: 181

### 한미약품 R&D센터

- 경기도 동탄
- 신약연구
- R&D 인력: 161



Total R&D staff  
(2024.3)

**642**



### 한미정밀화학 R&D센터

- 경기도 시화
- 원료의약품의 대량생산 연구
- R&D 인력: 62

### 제제 연구센터

- 경기도 팔탄
- 제제 연구, 약물전달기반 기술 연구
- R&D 인력: 58



### 바이오 공정 연구센터

- 경기도 평택
- 바이오공정 연구
- R&D 인력: 46

사업명		연합학습 기반 신약개발 가속화 프로젝트(K-MELLODDY) 데이터 활용 신약개발 협력 체계구축		
과제명		신약 개발 AI 모델의 정확도 향상을 위한 고품질 데이터 제공 및 최적화		
주관기관		한미약품(주)	연구책임자	최창주
전체 개발기간		2024년 7월 1일 ~ 2028년 12월 31일 (4년 6개월)		
최종 목표		FAM 구축을 위한 데이터 지원 및 활용으로 신약 개발 효율성 증대		
목표	1단계	기본 모델 개발을 위한 데이터 준비와 태스크 정의 및 데이터 공급 체계 구축	<ul style="list-style-type: none"> <li>• 데이터 보유 현황 조사 및 FAM 태스크 정의</li> <li>• 기본 모델 개발을 위한 데이터 제공 및 전처리 절차 수립</li> <li>• FDD 플랫폼 및 FAM 솔루션 구축을 위한 지원</li> <li>• 공급 데이터 품질 기준 정의 및 임상 데이터 활용 검토</li> <li>• 연합학습 플랫폼 데이터 공급</li> <li>• FDD 플랫폼 활용 및 FAM 솔루션 검증 지원</li> </ul>	
	2단계	데이터 공급 지속성 확보와 FDD 플랫폼 및 FAM 활용 검증을 통 한 ADMET 예측 활용	<ul style="list-style-type: none"> <li>• FDD 플랫폼 검증 및 데이터 지속 공급 계획 수립</li> <li>• AI 솔루션 성능 모니터링 및 활용 결과 분석</li> <li>• 적용 범위 확대 및 고도화 방안 연구</li> <li>• 학습 모델의 지속적인 검증 및 개선</li> </ul>	

## 초기 신물질 발굴 분야에서의 다양한 AI 활용 노력과 경험 보유



다수의 AI 활용 신약개발  
관련 정부 사업 참여(2018~)



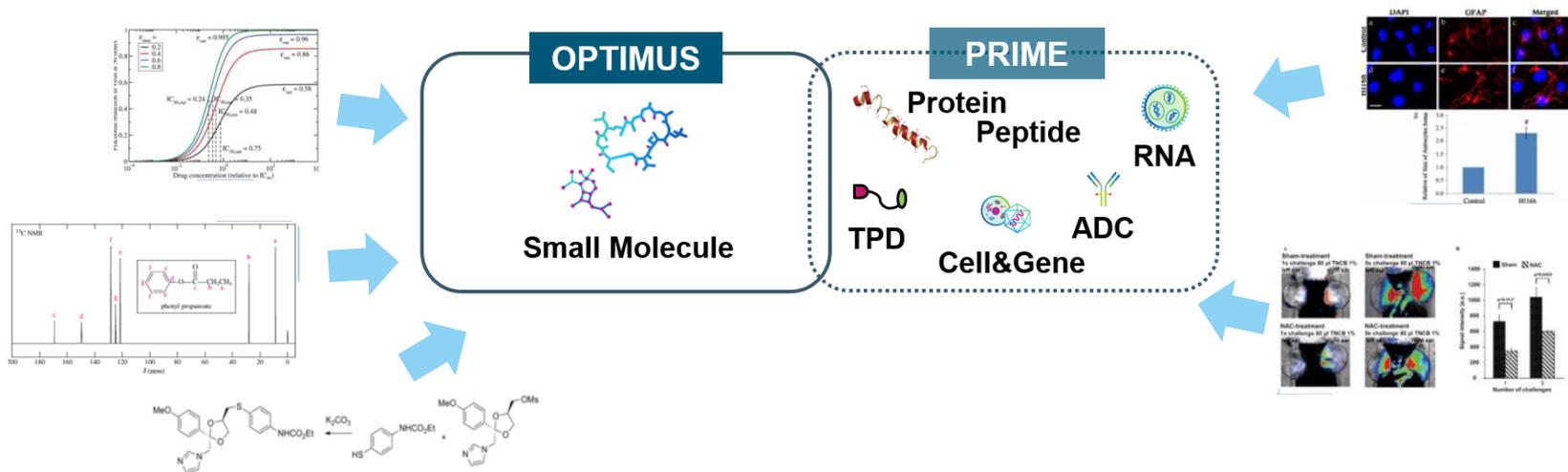
스탠다임, 신테카바이오,  
아이젠 사이언스와 공동연구



Digital Innovation TFT  
운영하여 내부 인프라 구축

## 모달리티별 체계적인 실험 데이터 관리 및 디지털화

각종 실험 데이터가 화합물 라이브러리로 수집



\*OPTIMUS (Optimized Precision Tools for Integrated Material Banks in a Unified System)

\*\*PRIME (Pharmaceutical Research Inventory and Management Engine)

## 품질 관리 체계를 통한 신뢰할 수 있는 데이터 보유

저분자 관련 50여개의 프로젝트, 11,000여종의  
고유 신물질 및 검증된 시험 데이터 보관

자체 표준작업지침서 준수로 모든 시험  
데이터의 무결성 보증을 위한 노력

**구조 정보**

- 화합물 코드 자동 부여
- ELN로부터 실험 결과물 연동
- Register에서 등록

**REGISTRATION INFORMATION**

Chemical Name	Registered Date	Chemistry	Type	Scientist
...	2021-12-01	...	Final product	System Admin

**STUDY RESULTS**

Experiment ID	Species	Route	Dose (mg/kg)	AUC(0-∞) (ng·h/mL)	Cmax or C3 (ng/mL)	Tmax (h)	CL (L/h/kg)	Vd (L/kg)	T1/2 (h)	SA (%)
167816	Rat	iv	3	850.2	441.7	...	2.6	18.4	5.6	...

**실험값 Assay Data 연동**

- In Vitro Assay
- Pharmacokinetics
- MS/PPB/PS
- CYP
- In Vivo Assay

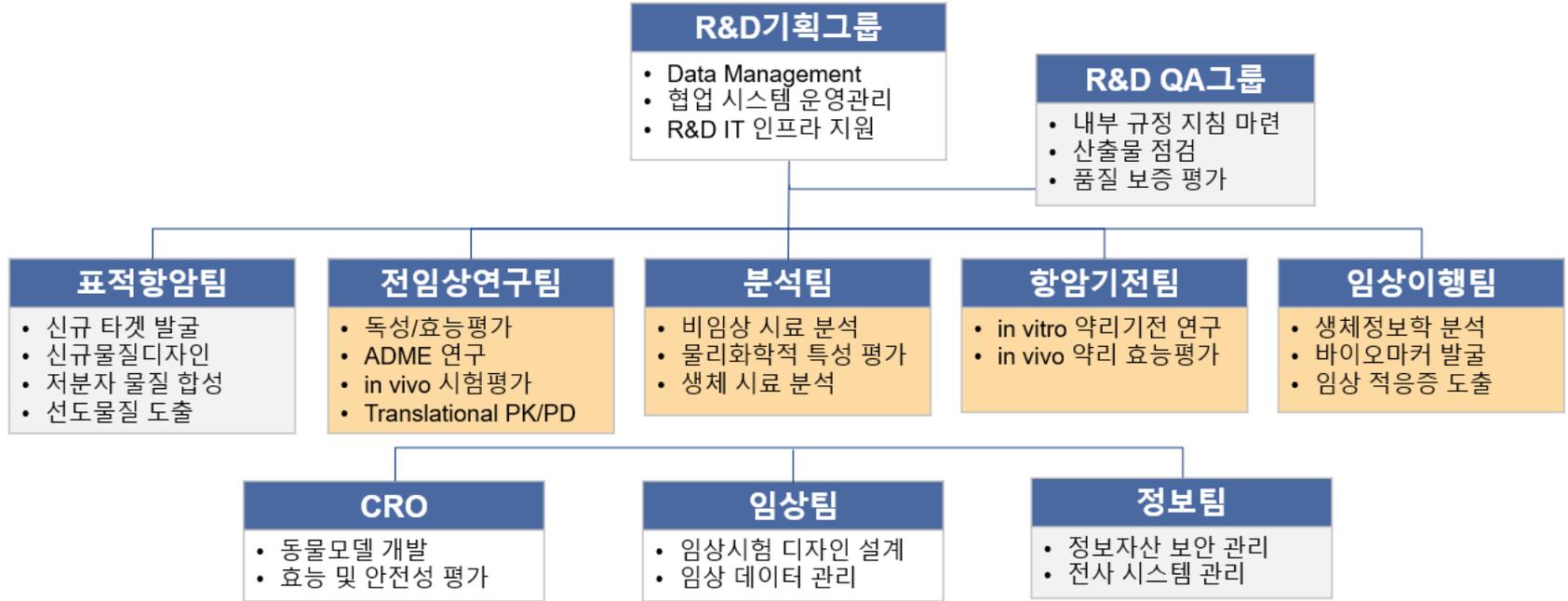
**Calculated Properties 자동 계산**

- XLogP, HBA, HBD
- Lipinski, RO3, FSP3
- TPSA, HAC, RotBondCount

XLogP	HBA	HBD	TPSA	HAC	RotBondCount
6.32	0	0	68.9	36	6



## 한미약품 내 유기적 유관 부서 협력 체계 및 주요 전략



### FAM 태스크 정의 방안

- 기존 ADMET 예측 도구 데이터 정리
- 신약 연구 성공 경험 공유, 유용한 예측 태스크 요청
- 각 조건별 시험에 대한 메타데이터 정보를 바탕으로 표준화 기준 마련

### 데이터 관리 및 품질 개선

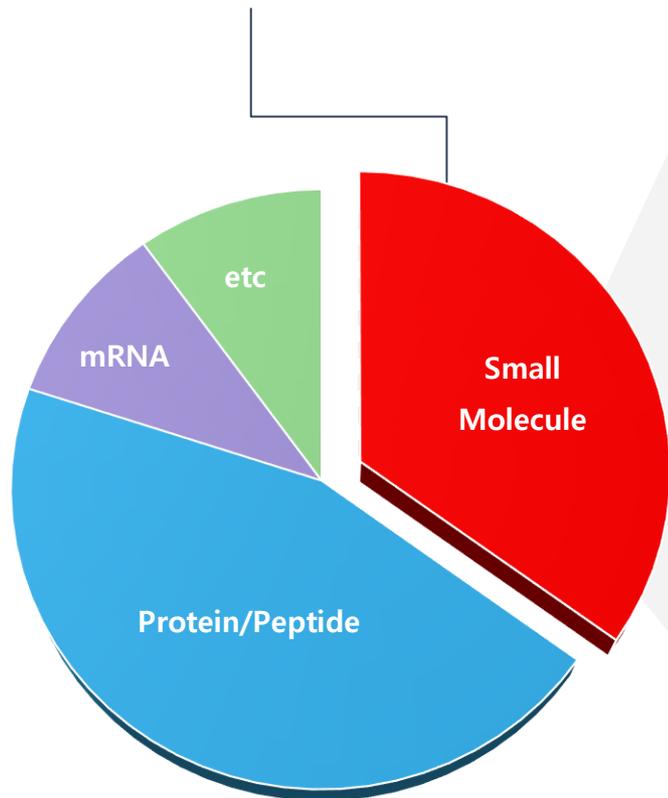
- ADMET 예측 모델의 각 태스크 정의에 따라 내부 관리 항목 개선
- 기존 자료 이관 및 지속적인 제공이 가능한 데이터 관리 방식 개선
- 품질 보증 점검으로 데이터 검증

### 지속적 데이터 공급 방안

- 지속적인 추가 실험 데이터 창출
- 기간별 항목 추출할 수 있도록 개선
- 다양한 생물학적, 화학적 데이터 소스 보강으로 예측 모델 정확도 개선
- 신약 임상 데이터의 적용 확대 검토

## 다양한 모달리티의 자체 IP 확보 물질 보유, 저분자 물질은 효력 데이터 위주

- 전체 물질의 약 40%가 **저분자 화합물 (11,000여종 보유)**
- 50여개의 과제, 100여종의 backbone 확보
- 매년 5~10개 신규과제 추가, 신규 물질 생성 증가
- \*북경한미 20여개 프로젝트의 화합물 데이터 추가 활용 가능**



분류	상세 항목	보유량
In Vitro	Cell/Enzyme Assay	19,709+
Physicochemical properties	용해도, 이온화 상수 (pKa), 친유성 분배계수 (XlogP)	93+
Absorption	Caco-2, PAMPA, 뇌장벽투과도	269+
	Mouse/rat/dog PK parameters	1,212+
Distribution	Plasma Protein Binding, Microsomal Binding in liver microsome	1,350+
	조직분포_분포용적	800+
Metabolism	Metabolic Stability, Plasma Stability	2,595+
	Metabolite identification	158+
Excretion	Clearance	800+
Drug-drug interaction	CYP inhibition assay, CYP induction assay, Phenotyping assay	636+
	Transporters substrate/ inhibitor	6+
Toxicity	hERG Fluorescence Polarization Assay, hERG assay (Patch-clamp)	349+
	Ames study, In vitro Mammalian Chromosomal Aberration assay, In vivo Micronucleus assay, In vitro Micronucleus assay, SOS chromotest	395+
	Seg I, Seg II, Seg III	11+
	정상세포 독성	65+
	In vitro phototoxicity, In vivo phototoxicity	8+

## 연합학습을 기반으로 국내 관련 기업들의 협업 문화 활성화 기대

- ✓ 기업 기밀의 약물 시험 데이터의 통합 활용으로 신약 발굴 효율 증대 및 제약 주권 확보 기대
- ✓ 내부 데이터 정제 및 가공 기술 이해와 표준화로 자체 데이터 축적 및 활용에 도움
- ✓ 각 기관 데이터의 보안 유지가 가능한 연합 학습 플랫폼의 타 산업분야 확장 가능성 제시

### Strengths

- R&D 중심 제약기업으로의 신약개발 역량
- AI를 활용한 신약개발에 대한 다양한 경험
- 내부 화합물 라이브러리의 시스템화
- 신뢰할 수 있는 데이터 품질 관리 체계

### Weaknesses

- 데이터가 특정 분야 (항암제, in vitro 평가)에 집중
- 임상 시험(in-human) 데이터 부족
- 저분자 ADMET 데이터 양적 부족
- 기관별 데이터 관리 기준 상이

### Opportunities

- 내부 데이터 표준화 및 정제
- 연합학습 예측 모델의 구축을 통한 새로운 도전
- 신약개발 R&D 투자 비용 절감
- AI 활용 신약개발 생태계 활성화

### Threats

- 저분자 이외의 다양한 모달리티 연구팀 확대
- 오류가 있는 데이터 제거의 어려움
- 빠르게 변화하는 데이터 분석 모델링 기법
- 다양한 상용 ADMET 예측 모델의 발전