

제1차 연합학습기반 신약개발 가속화 프로젝트사업 - 착수보고회

2-1 신약개발 데이터 활용 및 품질관리

KU Pharm 연합학습 기반 인공지능신약개발 데이터 랩

고려대학교 약학대학
이은희

목차

- 연구기관 및 연구자 소개
- 연구목표
- 연구의 핵심내용 및 기술
- 추진전략
- 추진체계
- 연구계획
- 향후 계획

연구 목표

Federated Drug Discovery 플랫폼 Federated ADMET Model 솔루션

신약개발 가속화
신약개발 성공률 향상
정보 보호와 데이터 활용 양립 기반구축

데이터 활용 신약개발 협력 체계구축
(KU Pharm 연합학습 기반 인공지능 신약개발 데이터 랩)

실험 Data
수집 및
가공

문헌 Data
수집 및
가공

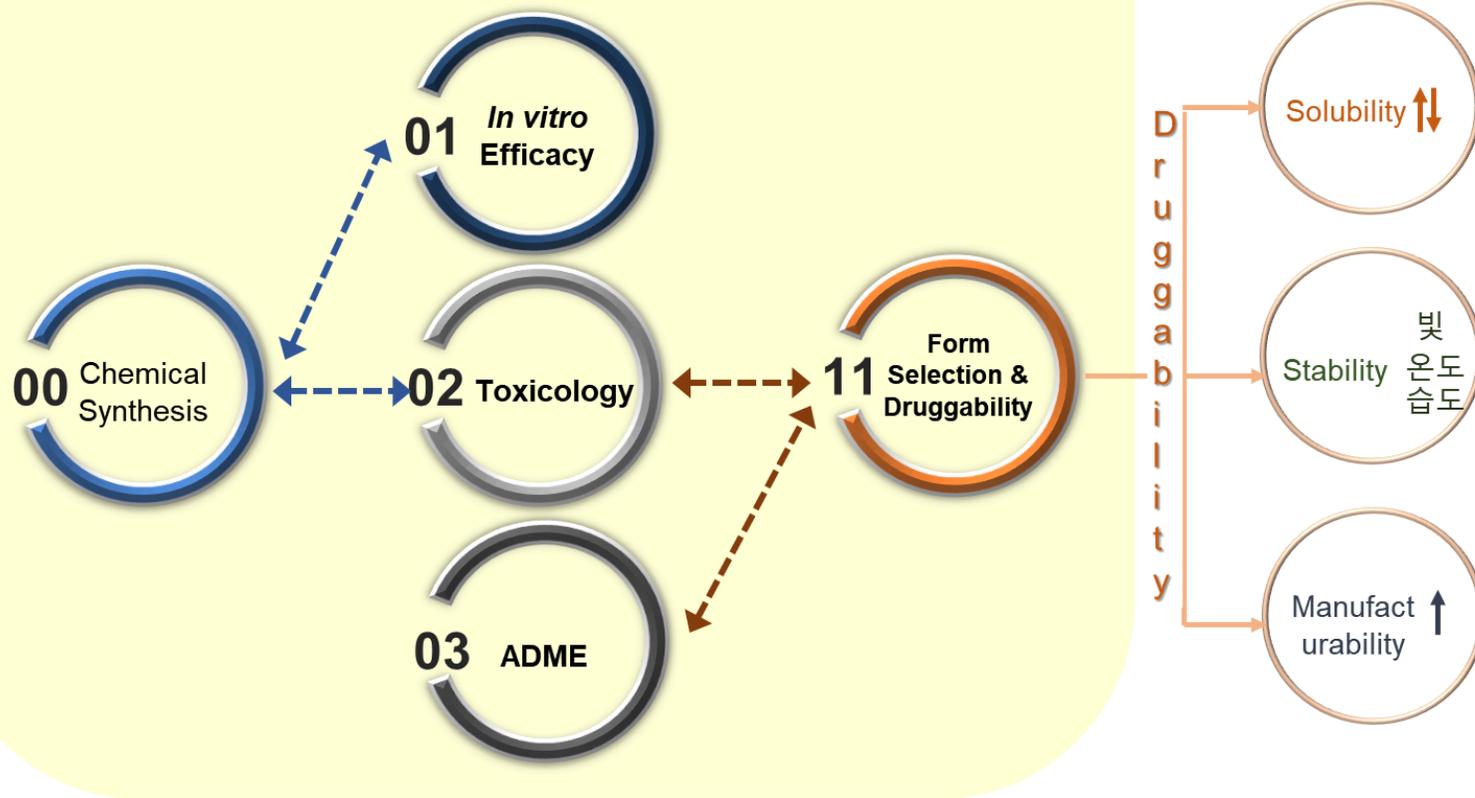
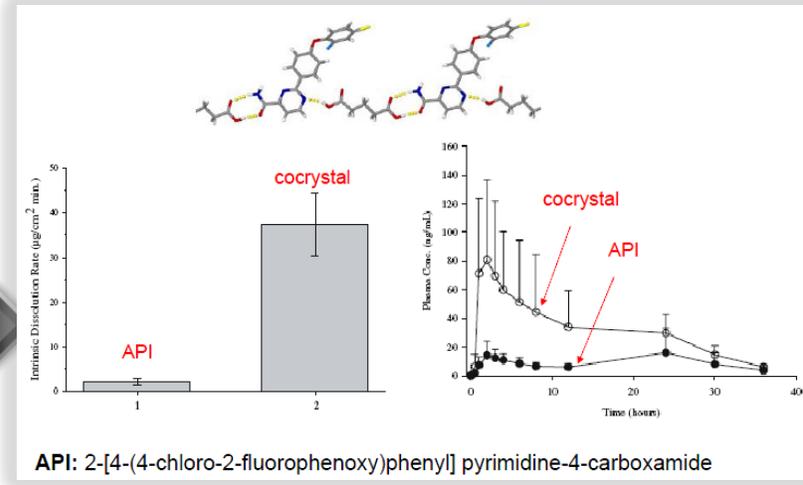
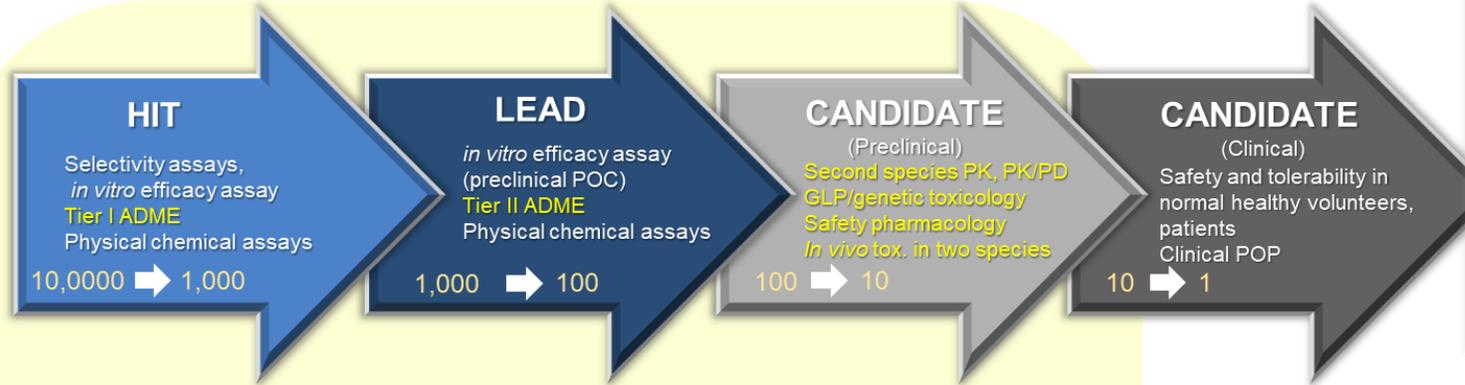
FAM 운영
성능개선
지원

FDD
플랫폼
확대 연구

신약개발 가속화, 신약개발 성공률 제고, 정보 보호와 데이터 활용 양립 기반구축을 위한 연합학습 기반 신약개발 (FDD, Federated Drug Discovery) 가속화 플랫폼 및 연합학습 기반 ADMET 예측모델(FAM, Federated ADMET Model) 솔루션 개발에 필요한

- 데이터를 생성하고,
- 활용할 수 있도록 전처리하며,
- 모델과 플랫폼의 활용 및 검증을 수행하여,
- 성능향상에 기여함

연구의 핵심내용



donepezil HCl tablets
5 mg
Rx only

CLOPIDOGREL TABLETS, USP
75 mg*
Rx Only

Atorvastatin Calcium Tablets
80 mg*
Rx Only

용해도

- Bigger and better databases containing "druglike molecules"
- Better measurement methods
- Sound experimental designed protocol
- Better 3D descriptors both in solution and describing the solid state

용해도: 정해진 온도와 압력하에, 정해진 용매에서 용질이 최대한 녹을 수 있는 양을 말하며, 이 때 녹아있는 용질은 **고체의 용질**과 동적 평형 상태임

고체의 성질: PXRD, DSC등을 활용한 고체상태 물성 파악



PXRD



TGA



DSC



Microscope

측정기기: visual, HPLC, UV 등

용매: DW, pH's, organic solvents

모델약물 선정: 프리폼, 염형, 공결정, 수화물, 용매화물 등



도네페질 염산염



암로디핀 베실산염



아토르바스타틴 칼슘삼수화물



클로피도그렐 황산염

매년 200개 이상의 데이터 생성 및 가공

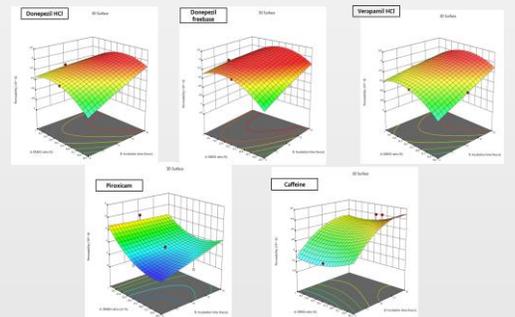
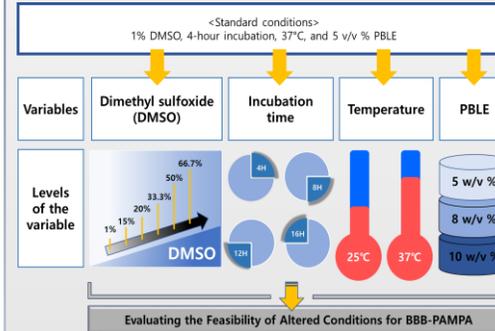
BBB-PAMPA

- Bigger and better databases containing "druglike molecules"
- Better measurement methods
- Sound experimental designed protocol
- Reconciliation of results from different measurement conditions

BBB-PAMPA: *in vivo* brain permeation data in rodents와 *in vitro* PAMPA-BBB data 예측률 77%

다양한 실험조건: 약물특성 (low solubility)을 반영한 실험조건

결과해석: QbD를 활용한 결과해석



BBB-PAMPA conditions 측정

용매: 약물의 용해도 고려, DMSO 비율 (1% - 66.7%)

온도: 25°C, 37°C

Incubation time: 4, 8, 12, 16 시간

PBLE 농도: 5, 8, 10 w/v %

측정: HPLC, LC-MS 등

모델약물 선정: 프리폼, 염형, 공결정, 수화물, 용매화물 등

매년 50개 이상의 데이터 생성 및 가공

보유데이터

문헌데이터 정리 및 가공

1. **FDA Orange Book Drug Lists**: 42,570개
(중복있음), CCDC에서 약물의 3D 구조
2. **유수저널 공결정 데이터 (500개)**: mw, mp, pKa, solubility 등 가능한 데이터 500개 이상 보유, 약물의 3D 구조 search

실험데이터 정리 및 가공

1. 용해도 데이터

- Cocrystal counterparts 40개 물 포함 12개의 유기용매에 대한 용해도 데이터 (480개)
- Salt counterparts 18개의 3개 용매에 대한 용해도 데이터 (54개)
- 허가된 약물 10여개의 다양한 용매에 대한 용해도 데이터 (134개)
- 20개 drug candidates의 23개 용매에 대한 용해도 데이터 (575개), confidential

2. BBB-PAMPA 데이터

- Caffeine, donepezil, donepezil HCl, piroxicam, verapamil HCl 5종 물질의 다양한 조건 (약 20개의 조건)에서 실험한 log Pe, membrane retention time data (총 200개)
- 약 8종 우울증 신약후보물질 BBB-PAMPA 조건에서 실험한 log Pe, MR (16개), confidential

보유 모델링 tools



CrystalExplorer

- ❖ **CrystalExplorer, freeware**
 - The Hirshfield surface calculation
 - Lattice energy calculation



The Cambridge Crystallographic Data Centre

❖ **CCDC, Mercury**

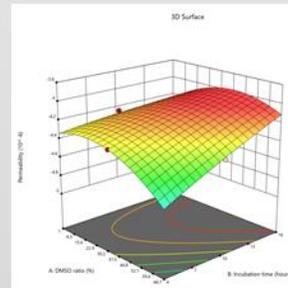
- 129만개 구조(2024. June)
- Knowledge-based approach
- Molecular synthon
- Cocrystal prediction
- Hydrogen-bond propensity



❖ **Materials Studio (BIOVIA)**

- 고체약물의 특성연구
- 결정구조 예측
- lattice energy calculation
- Quantum mechanical calculation
- Molecular mechanics calculation

DESIGN EXPERT



❖ **Quality by Design**

- Rational 실험디자인
- 실험조건별 결과해석
- 통계기법 적용

K-MELLODDY 사업단

FDD (Federated Drug Discovery) 플랫폼
FAM (Federated ADMET Model) 솔루션

실무위원회

운영협의체

연합학습 워크숍

KU Pharm 연합학습기반 인공지능신약개발 데이터 랩

1. 연합학습 플랫폼
구축 및 개발

FAM 모델 공급

2. 데이터 활용 신약개발
협력체계 구축

데이터 생성 및 전처리/
모델 성능 검증

2. 데이터 활용 신약개발
협력체계 구축

클라우드기반 데이터 축적
플랫폼 활용/검증

3. 연합학습 플랫폼
활용/활성화 지원

연합학습 플랫폼 구축 및 개발

• 데이터 전처리방법 제안

• FAM 모델 적합/필수적 데이터 및 포맷 제안

• 과제참여 커뮤니티 플랫폼 구축

• FAM 모델 공급

• 데이터 생성 및 전처리 수행

• FAM 모델 적합한/ 필수적인 실험, 문헌, modeling 데이터 생성

• 커뮤니티 활동참여 및 의견 개선

• FAM 모델 활용/검증, 피드백

• FDD 플랫폼 활용 및 피드백

• 클라우드기반 데이터 축적 및 활용

• 사용자 인터페이스 활용 및 피드백

• FDD 활용 후 피드백, 개선사항 제안

• FDD 플랫폼 구축 및 개선

• 클라우드기반 데이터 관리 협력

• 단계별 플랫폼 구축 전략 제안

• 연합학습 플랫폼 FDD 요구사항 정의/설계/개발

WORKING GROUP
사용자 편이성, 수요자맞춤형

연구 계획 및 산출물

실험약물
수집 및
보유데이터
전처리

실험 Data
수집/가공
협력

문헌 Data
수집/가공
협력

Modeling
/Exp.
Hybrid

모델/
플랫폼
검증/고도화

- 보유데이터 조사 및 가공 – 문헌 데이터
 - FDA approved drug lists 확보 → 단결정 구조 찾기 (CCDC database search), 실험약물 lists 우선순위
 - 논문데이터 (약 500개의 우수 논문의 공결정 구조 및 용해도 데이터)
- 보유데이터 조사 및 가공 – 실험 데이터
 - 20여개의 신약후보물질 solubility (water, 32개 유기용매, pH 1-12) 및 일부 PK data, *confidential data*
 - 10여종의 허가된 약물 solubility (water, 32개 유기용매, pH 1-12)
 - 약물 및 유기물질 물에 대한 용해도 data
 - 다양한 실험조건 별 BBB-PAMPA data 약 100개 확보, 신약후보물질 BBB-PAMPA data 확보
- 실험약물 수집 및 데이터 생성: 매년 200개 이상의 solubility data, 매년 50개 이상의 BBB-PAMPA data

- 문헌/특허: 우수저널 데이터 및 SciFinder 활용, 1세부 협력 데이터 전처리
- DB 활용: 10,000,000개 이상의 crystal structure database (CSD), 42,570개 이상의 FDA Orange Book Drug의 physicochemical property 전처리, 한국화학연구원 화합물 은행 데이터 서치
- 연합학습 상호운용성을 위한 데이터 준비 및 저장 (수집/정리/정제/가공 등 전처리 작업 수행)

- Crystal structure database (CSD) 활용 descriptor 개발 협력
- Materials Studio (BIOVIA) property prediction 활용 교차검증
- Modelling 기법: Knowledge-based approach, QM/MM 활용, COSMO 활용
- 다양한 모델링 기법/ QbD 방법과 ML 기법의 비교분석

- 클라우드에 데이터 축적 및 활용
- FDD를 통한 FAM의 기관별 로컬 학습 수행 및 성능 모니터링 및 활용
- FDD 플랫폼 및 FAM 솔루션 활용 및 검증, 피드백 제공
- FDD 플랫폼 적용 범위 확대 및 FAM 활용 방안 연구
- Virtual Salt/Cocrystal Screening Tool 개발을 위한 협력
- 세미나/심포지엄/협의체 참여 및 의견개진, 데이터 고도화 방안 논의

- 보유데이터 전처리
 - FDA approved drug lists 확보
- 용해도 데이터 생성 및 전처리
 - 매년 200개 이상 생성
 - 생성된 데이터 전처리
- BBB-PAMPA 데이터 생성 및 전처리
 - 매년 50개 이상 생성 및 전처리

- 문헌데이터 정리 및 전처리
 - 매년 200개 이상 정리 및 전처리
 - 다양한 데이터 전처리
- 각종 database 서치 및 데이터 전처리
 - 매년 200개 이상 정리 및 전처리

- Modeling 기법 활용 교차검증
 - CSD 활용 descriptor 정리
 - Modeling 기법 활용 교차검증

- FDD 플랫폼/FAM 솔루션 활용 및 검증
 - 활용보고서 (2단계)
 - 분석결과보고서 (2단계)
 - 검증보고서 (2단계)
 - 참여실적